Option Pricing for Alternative Risky Asset Price Models, and Model Verification Results

A thesis submitted for the degree of Doctor of Philosophy of

The Australian National University

Benjamin David Kaehler

August 2009

To Mum and Dad

Declaration

The work in this thesis is my own except where otherwise stated.

Ben Kaehler

Acknowledgements

I consider myself extremely privileged to have received the support of my supervisors. Thank you Chris Heyde, Ross Maller, and Alex Szimayer.

Thanks are due to my fellow students Allan Sly, Damien Bankovsky, Pourus Bharucha, Griff Ware, and Michael Roper for their camaraderie and for teaching me mathematics, and especially to Allan. I am grateful to the Fraunhofer ITWM Finanzmathematik for their hospitality and in particular Sascha Desmettre, Johan De Kock, Christina Erlwein, and Peter Ruckdeschel. Thanks go to David Heath and Timo Vollmer for collaboration, to Vladislav Kargin for correspondence, to John Trueman for meticulous proofreading, and to Wendy Anderson.

I also appreciate the boundless and constant support of my family, thank you Holly, Julie, Harry, Jack, Cleo, Will, Mum and Dad.

Abstract

We combine two aspects of modern option pricing, where risky assets are not necessarily modelled using Geometric Brownian Motion. The first focuses on pricing American rainbow options using Lévy models; the second addresses model verification where the log returns of price processes are allowed to display both leptokurtosis and long-range dependence.

The calculation of American option prices under the assumption that the price process follows an exponential Lévy process requires specialised numerical techniques. The computation of American prices for rainbow options, also known as multi-asset options, also presents challenges. Motivated by a multinomial tree model for pricing exotic options using Lévy processes and a novel tree-based approach to pricing high-dimensional rainbows under Black-Scholes assumptions, a discretisation scheme for any multivariate Lévy process is presented, with corresponding results regarding the convergence of American prices. Further, a new, general result pertaining to the convergence of Bermudan prices to American prices for one-dimensional Lévy models is given. The practical implications of these results are demonstrated with the implementation of a simple pricing algorithm.

Many modern pricing models allow risky asset log returns to be skewed, but skewness can be difficult to measure directly, especially when samples are not necessarily independent or mesokurtic. We eschew the usual assumption of independent identically distributed samples in favour of the looser assumption of stationary ergodic martingale differences and present a class of generalised skewness statistics. This class reduces to the usual skewness statistic when it is applicable, but may also be used when a sample's distribution is believed to be leptokurtic. We show how the statistics can be studentised for use in hypothesis testing by investigating their asymptotic distributions. The new statistics are applied to a large data set from the Australian share market.

Contents

Chapter 1

Introduction

It is difficult to discuss option pricing without couching the discussion in the context of the seminal work of Black and Scholes (1973). This thesis is concerned with the models for risky asset prices that come after those presented in Black and Scholes (1973). Part I is about option pricing in markets where risky assets are assumed to follow exponential Lévy processes. This represents a set of market models that contains the model assumed in Black and Scholes (1973). Part II looks at how we might verify statistical claims about risky asset returns if we do not assume a Black-Scholes model, with implications for the formulation of new option pricing models. A common theoretical thread for Parts I and II is that they are mainly involved with asymptotic results pertaining to stochastic processes.

Within the field of pricing options assuming Lévy models, Part I addresses the topic of pricing American rainbow options using lattice, or tree-based, methods. A rainbow option is any derivative where the payoff is a function of more than one risky asset price. In particular, we seek to generalise the multinomial tree model introduced in Maller, Solomon, and Szimayer (2006), which can price American options on single assets for any Lévy model with a known Lévy measure.

This generalisation faces problems theoretical and practical. The theoretical problems are related to the availability of results pertaining to optimal stopping

problems in multiple dimensions, referring in particular to the results in Mulinacci and Pratelli (1998) that are used in Maller et al. (2006). These challenges were overcome and our main result is the option price convergence result given in Theorem 3.4.2. The practical difficulty is the increased computational burden of performing a lattice based calculation in higher dimensions. This is an example of what is commonly known as the curse of dimensionality. Motivated by an algorithm for pricing American options in a highly multidimensional Black-Scholes setting presented in Kargin (2005), we modify the model presented in Maller et al. (2006) to highlight to a potential user of the model the efficiency trade-offs associated with implementing it numerically.

An outcome of this generalisation is the discovery of a result that is new in both one and multiple dimensions. A feature of the multinomial model described in Maller et al. (2006) is that it is possible to prove that approximate option prices obtained with this model converge asymptotically to the corresponding continuous time option prices, as is shown in Szimayer and Maller (2007). Proof of this convergence has not previously been obtained for other multinomial models for Lévy processes present in the literature. For instance, in Këllezi and Webber (2004), four multinomial models for pricing Bermudan options using Lévy processes are presented, but the authors have difficulty demonstrating numerical convergence of Bermudan option prices to American option prices with increasing numbers of exercise opportunities. Our result, Theorem 3.4.1, proves convergence of a Bermudan option price to the corresponding American option price as the number of exercise opportunities increases. This proof is valid for Lévy models in one dimension with mild constraints on the payoff function. This development allows us to show that the prices obtained using some other multinomial models do converge, or how these models can be modified so that they will. See Remarks 2.3.1, 2.4.1, and 2.4.2 and Section 3.5 for more details. We also generalise Theorem 3.4.1 for rainbows, although with a weaker mode of convergence, in Theorem 3.4.3.

Chapter 2 provides a literature review with some new results. It introduces the market model that is assumed throughout and gives some particular examples of Lévy processes. The new results include giving important Lévy measures and transition probability density functions for two dimensional Lévy processes in closed form. We then explore how lattice models fit into the technology available for pricing American options using Lévy processes, in one and several dimensions. We also introduce the theoretical context required for developments in Chapter 3.

Chapter 3 states and proves our convergence results. We first set up two approximation schemes, the first jump approximation model and the skeleton approximation model. The first jump approximation model is the generalisation of the multinomial model introduced in Maller et al. (2006). The skeleton approximation model allows us to prove the Bermudan convergence result in Theorem 3.4.1. After proving certain tightness and convergence results for these approximation models, we generalise some theories from Mulinacci and Pratelli (1998) to the multidimensional setting, allowing us to prove the rainbow price convergence results in Theorems 3.4.2 and 3.4.3. Chapter 3 also states several related conjectures and provides an example of an existing model from Këllezi and Webber (2004) that can be analysed in the light of Theorem 3.4.1.

Chapter 4 demonstrates how the approximation schemes developed in Chapter 3 can be used in practice by way of an example. It presents a general algorithm, based on the work in Kargin (2005), and an example software architecture that can in principal be implemented to evaluate the price of any American option with a continuous bounded payoff using any Lévy model with either the skeleton or first jump approximation models. It then gives a concrete example by way of valuing a particular American option assuming a specific Lévy model. The results of some numerical experiments are given and the implications for more industrial implementations are discussed. Chapter 5 concludes Part I.

The presence of skewness in financial time series is of great importance in the

fields of asset allocation and option pricing, where information about the asymmetry of returns can potentially be used to create positively skewed returns on stock portfolios (Konno, Shirakawa, and Yamazaki 1993), or to explain the so called *volatility smirk* observed in options markets (Hull 2003, pp. 438–440). The volatility smirk is an example of a *stylised feature* that is not consistent with the Black-Scholes model. Almost all of the Lévy models that are reviewed in Chapter 2 allow the log returns of risky assets to be skewed. Whether or not risky asset log returns are skewed then is a pertinent question when selecting risky asset models, and this question is addressed in Part II in the context of a class of risky asset models that contains many Lévy models, as well as some that are not.

Testing skewness in financial data is problematic, especially when the test statistic involves the empirical third moment of the sample. In particular, considerable difficulties can arise when the distribution from which the sample is drawn is leptokurtic. Further, when deriving properties of such statistics it is usually assumed that the samples are identically and independently distributed (iid), an assumption that is not born out in some observations of financial time series.

Chapter 6 defines a class of skewness statistics that generalises the usual sample third moment based skewness statistic and derives the statistics' asymptotic distributions, assuming only that the samples form a series of stationary ergodic martingale differences with finite variance, which is to say that stocks obey the efficient market hypothesis and have finite volatility. The main result is that the new statistics are shown, when studentised, to tend in distribution to the standard normal distribution for large samples, under mild conditions. This result, though to be expected, perhaps, requires some quite intricate analysis for its proof, and the asymptotic variance, needed for the studentisation, is not trivial to calculate. We also test the results of Chapter 6 using simulation.

Chapter 7 reviews the financial setting for our proposed skewness statistics and some of the efforts of others who have attempted to measure skewness of financial

time series directly. We then apply the new skewness statistics to a broad set of Australian share price time series, in an attempt to assess the extent of skewness of the distributions of their daily log returns. Part II is concluded in Chapter 8.

Part I

Pricing American Rainbow Options Using Lévy Processes

Part I

Pricing American Rainbow Options Using Lévy Processes

Chapter 2

Context

2.1 Lévy Processes and American Prices

Black and Scholes (1973) presented an elegant model for pricing options and had an impact on markets globally that cannot be overstated. For a readable reference on derivatives markets under Black-Scholes assumptions see Hull (2003). As for any model, however, there are discrepancies between the predictions of the Black-Scholes model and reality. Some discrepancies are due to technical assumptions such as frictionless markets. Others are caused by more fundamental assumptions, such as that the log returns on risky assets are normally and independently distributed, or the implicit assumption that markets are complete, that is that the risk associated with any contingent claim can be perfectly hedged using a dynamic portfolio of tradeable assets. For a survey of the ways in which the assumption of independent normally distributed log returns is violated by real assets see Rydberg (2000).

In the pursuit of more consistent option prices, more sophisticated models have been developed. Many fall into the class of Lévy models, in which risky asset prices are represented by exponential Lévy processes. While the Black-Scholes model is in this class, no other Lévy model assumes log normally distributed returns, and for most Lévy models the market is not complete. Cont and Tankov (2004) is a good

reference for the use of Lévy models in finance. Part I of this thesis develops results that are applicable to Lévy models in general.

On a probability space $(\Omega, \mathbb{F}, \mathrm{P})$, a Lévy process $L(t)$ is defined to be a stochastic process on $\mathbb{R}^d$ that has independent, stationary increments with $\|L(t+h)-L(t)\| \xrightarrow{\mathrm{P}} 0$ as $h \to 0$ for all t, where $\|\cdot\|$ denotes the Euclidean norm. More intuitively, a Lévy process can be defined as the pathwise sum of a Wiener process and a possibly infinite number of independent Poisson processes (Cont and Tankov 2004, Chapter 3). $L(t)$ can be characterised by its Lévy triplet (γ, A, Π), where $\gamma \in \mathbb{R}^d$, A is a symmetric nonnegative-definite $d \times d$ matrix, and Π is a measure on $\mathbb{R}^d$ satisfying $\Pi(\{0\}) = 0$ and $\int_{\mathbb{R}^d} (\|x\|^2 \wedge 1)\, \Pi(\mathrm{d}x) < \infty$. A and Π are known as the Gaussian covariance matrix and the Lévy measure respectively. Let $\langle \cdot, \cdot \rangle$ be the scalar product. Define $D = \{x : \|x\| \leq 1\}$ and write the canonical exponent of $L(t)$, which satisfies $\mathrm{E}\, \mathrm{e}^{i\langle \theta, L(t) \rangle} = \mathrm{e}^{-t\Psi(\theta)}$, as

$$\Psi(\theta) = \frac{1}{2}\langle \theta, A\theta \rangle - \mathrm{i}\langle \gamma, \theta \rangle + \int_{\mathbb{R}^d} \left(1 - \mathrm{e}^{\mathrm{i}\langle \theta, x \rangle} + \mathrm{i}\langle \theta, x \rangle \, \mathbf{1}_D(x)\right) \Pi(\mathrm{d}x),$$

where $\theta \in \mathbb{R}^d$. These specific properties can be found at Sato (1999, p. 37). Sato (1999) is a very useful reference for Lévy processes. See Cont and Tankov (2004, Chapter 3) for a concise and not overly technical overview. For probabilistic underpinnings see Billingsley (1995) and Billingsley (1999). Section 2.2 will give several examples of Lévy processes relevant to this work.

Explicitly, we assume a market model that contains a risk free asset with price e^{rt} at time t and d risky assets modelled as

$$S_j(t) = S_j(0)\mathrm{e}^{L_j(t)}, \quad j \in \{1, \ldots, d\}, \quad t \in [0, T], \tag{2.1.1}$$

where $L_j(t)$ are the d components of the Lévy process $L(t)$, and $S_j(t)$ are the d components of $S(t)$ on $\mathbb{R}^{+d}$. $S(0)$ is the non-stochastic starting price. Let $\mathbb{F}^L = (\mathcal{F}_t)_{t \geq 0}$ be the natural filtration of $L(t)$. We also assume that $\mathrm{e}^{-rt} S_j(t) = \mathrm{E}\left(\mathrm{e}^{-rT}\, S_j(T) | \, \mathcal{F}_t\right)$

for $j \in \{1, \ldots, d\}$ and $0 \leq t \leq T$ and that P is equivalent to the physical measure. Thus, P is an equivalent martingale measure, and the market model is arbitrage-free (Cont and Tankov 2004, Section 9.1.1). The method by which this measure might be found is outside the scope of this work. Calibration to real markets is covered in Cont and Tankov (2004, Chapter 13).

An American option can be exercised at any time up to the maturity of the option at time T, at which point the holder receives a payoff $h : [0,T] \times \mathbb{R}^{+d} \mapsto \mathbb{R}^+$, dependent on the prices of the d risky assets. For any scenario where $d > 1$, the option can be termed a rainbow option. Rainbow options will be discussed in Section 2.4. If the option has not been exercised before time t, the price process of the option $\pi(t)$ can be written as the solution to the optimal stopping problem

$$\pi(t) = \operatorname*{ess\,sup}_{\tau \in \mathcal{S}_{t,T}} \mathrm{E}\left(\mathrm{e}^{-r(\tau-t)} h(\tau, S(\tau)) \middle| \mathcal{F}_t\right), \quad t \in [0,T], \tag{2.1.2}$$

where $\mathcal{S}_{t,T}$ is the set of $\mathbb{F}^L$-stopping times in $[t,T]$ (Cont and Tankov 2004, Section 11.4). For given stock prices at time t, $x \in \mathbb{R}^{+d}$, the function v such that $v(t, S(t)) = \pi(t)$ reduces to

$$v(t,x) = \sup_{\tau \in \mathcal{S}_{0,T-t}} \mathrm{E}\left(\mathrm{e}^{-r\tau} h\left(t+\tau, x_1 \mathrm{e}^{L_1(\tau)}, \ldots, x_d \mathrm{e}^{L_d(\tau)}\right)\right). \tag{2.1.3}$$

Numerical methods for calculating $v(t,x)$ for some specific Lévy processes will be reviewed in Section 2.3.

2.2 Lévy Process Examples

The class of Lévy models is very broad. We mention several examples here for future reference. For a more comprehensive list with greater depth of discussion see Cont and Tankov (2004, Chapters 3 and 4). Some one-dimensional models are given before multidimensional Lévy processes are addressed. Some results regarding

multidimensional processes that do not appear in the literature are presented.

2.2.1 One Dimensional Lévy Processes

Two fundamental examples of Lévy processes are Gaussian processes and Poisson processes. One dimensional Gaussian and Poisson processes have Lévy triplets of the form $(\gamma, A, 0)$ and $(0, 0, \Pi)$ respectively, where $\Pi(\mathrm{d}x) = \lambda\delta_0(x-1)\mathrm{d}x$ for Poisson intensity λ and δ_0 is the Dirac delta measure. Multidimensional Gaussian processes have Lévy triplets of the same form, where A becomes the variance-covariance matrix. Lévy processes can be divided neatly into jump-diffusion type processes and processes with *infinite activity* (Cont and Tankov 2004, Section 4.1.1). Jump-diffusion processes can be expressed in the form

$$L(t) = \gamma t + \sigma W(t) + \sum_{i=1}^{N(t)} X(i),$$

where $W(t)$ is a Wiener process, $N(t)$ is an independent Poisson process of intensity λ, and $\{X(i)\}_{i\in\mathbb{N}}$ is a sequence of iid random jumps with probability distribution function f. The Lévy triplet of such a process is (γ, σ^2, Π), where $\Pi(\mathrm{d}x) = \lambda f(x)\,\mathrm{d}x$ (Cont and Tankov 2004, Section 4.3). Several types of jump-diffusion processes exist in the literature, differentiated by their jump size probability distributions. Processes of infinite activity cannot be expressed in this fashion as the Lévy measure of such processes has a singularity at zero. That is, they have an infinite number of small jumps in any finite time interval. Infinite activity processes provide a rich class of models, and can sometimes be constructed via the subordination of a Gaussian process.

Any (multidimensional) Lévy process subordinated by an independent Lévy subordinator is also a Lévy process (Cont and Tankov 2004, Theorem 4.2). A Lévy subordinator is a Lévy process with almost surely increasing paths, or for equivalent definitions see Theorem 3.10 in Cont and Tankov (2004). In finance the

intuition behind subordination is to say that the evolution of a risky asset price occurs at some rate that is not clock time, but rather the rate at which the market is moving, some *business time*. We introduce two Lévy subordinators, the Gamma process and the Inverse Gaussian process (Cont and Tankov 2004, Section 4.4.2). A Lévy process can be specified by giving its marginal (infinitely divisible) distribution at any fixed time. The Gamma process $G(t)$ is defined so that $G(1) \sim \Gamma(k, \lambda)$. That is, $G(1)$ is Gamma distributed with probability density function $f(x) = \lambda^k x^{k-1} \mathrm{e}^{-\lambda x} / \Gamma(k)$ for some $k, \lambda > 0$ and all $x > 0$. The Lévy measure for the Gamma process is given by $\Pi(\mathrm{d}x) = k\mathrm{e}^{-\lambda x} x^{-1} \mathrm{d}x$. Subordinators are sometimes described so that their expectations grow at the same rate as clock time, that is so that $\mathrm{E}\, G(t) = t$. To achieve this condition for the Gamma process with the above parameters we set $\lambda = k$. Similarly, the Inverse Gamma process $I(t)$ is defined so that $I(1) \sim \mathrm{IG}(k, \lambda)$ with the Inverse Gamma probability density function defined as $f(x) = kx^{-\frac{3}{2}} \exp\left\{ -\left(\sqrt{\lambda}x - k\sqrt{\pi}\right)^2 \Big/ x \right\}$ for $k, \lambda > 0$ and all $x > 0$, and Lévy measure $\Pi(\mathrm{d}x) = k\mathrm{e}^{-\lambda x} x^{-\frac{3}{2}} \mathrm{d}x$. We can ensure that $\mathrm{E}\, I(t) = t$ in this context by setting $\lambda = k^2 \pi$.

The Variance Gamma (VG) process is obtained by subordinating a Gaussian process with a Gamma process, and was introduced for the purpose of modelling stock returns in Madan and Seneta (1990). It is a pure jump process. If we take a Gamma process $G(t)$ such that $G(1) \sim \Gamma(b, b)$, $b > 0$, and a Brownian process $B(t)$ with drift $\mu \in \mathbb{R}$ and volatility $\sigma > 0$, the transition probability density function and Lévy measure of the VG process $L(t) = B(G(t))$ are given respectively by (Madan, Carr, and Chang 1998)

$$f_t(x) = \frac{2b^{bt}}{\sqrt{2\pi}\sigma\Gamma(bt)} \left(\frac{x^2}{\mu^2 + 2b\sigma^2} \right)^{\frac{bt}{2} - \frac{1}{4}} \exp\left\{ \frac{\mu x}{\sigma^2} \right\} \mathrm{K}_{bt - \frac{1}{2}} \left(\left(\frac{\mu^2}{\sigma^2} + 2b \right)^{\frac{1}{2}} \frac{|x|}{\sigma} \right)$$

$$\text{and} \quad \Pi(\mathrm{d}x) = \frac{b}{|x|} \exp\left\{ \frac{\mu x}{\sigma^2} - \left(\frac{\mu^2}{\sigma^2} + 2b \right)^{\frac{1}{2}} \frac{|x|}{\sigma} \right\} \mathrm{d}x, \tag{2.2.1}$$

where $\mathrm{K}_\nu(z)$ is the modified Bessel function of the second kind (Abramowitz and Stegun 1964, p. 374). It is clear from its Lévy measure that the VG process is an infinite activity process.

The Normal Inverse Gaussian (NIG) process is similarly generated by subordinating a Gaussian process with an Inverse Gaussian process, and is also a pure jump process. See, for example, Barndorff-Nielsen (1997). If we take a Gaussian process $B(t)$ as defined in the last paragraph and an Inverse Gamma process $I(t)$ such that $I(1) \sim \mathrm{IG}\left(\frac{b}{\sqrt{2\pi}}, \frac{b^2}{2}\right)$, the transition probability density function of the NIG process $L(t) = B(I(t))$ is given by (Cont and Tankov 2004, Section 4.4.3)

$$f_t(x) = \frac{b}{\pi\sigma} t \left(\frac{\mu^2 + b^2\sigma^2}{x^2 + b^2\sigma^2 t^2}\right)^{\frac{1}{2}} \exp\left\{\frac{\mu}{\sigma^2}x + b^2 t\right\} \mathrm{K}_1\left(\frac{1}{\sigma^2}\left(\mu^2 + b^2\sigma^2\right)^{\frac{1}{2}}\left(x^2 + b^2\sigma^2 t^2\right)^{\frac{1}{2}}\right),$$

and the Lévy measure is

$$\Pi(\mathrm{d}x) = \frac{b}{\pi\sigma}\left(\mu^2 + b^2\sigma^2\right)^{\frac{1}{2}} \frac{1}{|x|} \exp\left\{\frac{\mu}{\sigma^2}x\right\} \mathrm{K}_1\left(\frac{1}{\sigma^2}\left(\mu^2 + b^2\sigma^2\right)^{\frac{1}{2}}|x|\right) \mathrm{d}x.$$

As for the VG process, the NIG process possesses infinite activity.

The last two processes were described using three parameters each. Another pure jump, infinite activity time process, the CGMY process, which has four parameters, $C > 0$, $G \geq 0$, $M \geq 0$, and $Y < 2$, was introduced in Carr, Geman, Madan, and Yor (2002). Its Lévy measure is given by

$$\Pi(\mathrm{d}x) = \left[\frac{C\mathrm{e}^{Gx}}{(-x)^{1+Y}}\mathbf{1}_{x<0} + \frac{C\mathrm{e}^{-Mx}}{x^{1+Y}}\mathbf{1}_{x>0}\right]\mathrm{d}x.$$

The CGMY process is an example of a *tempered stable process* (Cont and Tankov 2004, Remark 4.3), and in general there is no closed form expression for its transition probability density function.

2.2.2 Multidimensional Lévy Processes

There are several ways in which one might define a multidimensional Lévy process. We do not attempt an exhaustive review of the subject, or make judgements on the relative merits of the different approaches, but give some examples that may be useful in the context of pricing rainbow options. Note that all of the methods presented in the following examples, which mostly happen to be two-dimensional, can be generalised to higher dimensions.

It is mentioned in Madan and Seneta (1990) that the VG process can be generalised to arbitrary dimensions by subordinating a multivariate Brownian process by the same subordinator. That is, if we take a Gamma process $G(t)$ such that $G(1) \sim \Gamma(b, b)$ and a Brownian process $B(t)$ in d dimensions with mean vector μ and covariance matrix Σ, a multivariate VG process can be formulated as

$$L(t) = B(G(t)). \tag{2.2.2}$$

The transition probability density function and the Lévy measure for the bivariate case are given in the following theorem. This result does not appear to be present in the literature. Note that the form of these equations changes with dimensionality. Compare the below expressions with (2.2.1).

Theorem 2.2.1. *Take the process defined in* (2.2.2) *with* $d = 2$. *Its transition probability density function and Lévy measure are*

$$\begin{aligned} f_t(x) = {} & \frac{b^{bt}}{\pi \det(\Sigma)^{\frac{1}{2}} \Gamma(bt)} \exp\left\{\left\langle \mu, \Sigma^{-1} x \right\rangle\right\} \\ & \times \left[\left(\frac{\langle x, \Sigma^{-1} x \rangle}{\langle \mu, \Sigma^{-1} \mu \rangle + 2b} \right)^{\frac{bt}{2} - \frac{1}{2}} \mathrm{K}_{bt+1}\left(\left(\left\langle \mu, \Sigma^{-1} \mu \right\rangle + 2b \right)^{\frac{1}{2}} \left\langle x, \Sigma^{-1} x \right\rangle^{\frac{1}{2}} \right) \right. \\ & \left. - \frac{2bt \left\langle x, \Sigma^{-1} x \right\rangle^{\frac{bt}{2} - 1}}{\left(\left\langle \mu, \Sigma^{-1} \mu \right\rangle + 2b \right)^{\frac{bt}{2} - 2}} \mathrm{K}_{bt}\left(\left(\left\langle \mu, \Sigma^{-1} \mu \right\rangle + 2b \right)^{\frac{1}{2}} \left\langle x, \Sigma^{-1} x \right\rangle^{\frac{1}{2}} \right) \right] \end{aligned}$$

and

$$\Pi(\mathrm{d}x) = \frac{b}{\pi}\left(\frac{\langle\mu,\Sigma^{-1}\mu\rangle + 2b}{\det(\Sigma)\,\langle x,\Sigma^{-1}x\rangle}\right)^{\frac{1}{2}} \exp\left\{\langle\mu,\Sigma^{-1}x\rangle\right\} \times \mathrm{K}_1\left(\left(\langle\mu,\Sigma^{-1}\mu\rangle + 2b\right)^{\frac{1}{2}}\langle x,\Sigma^{-1}x\rangle^{\frac{1}{2}}\right)\mathrm{d}x.$$

This may be proved with the assistance of the following lemma.

Lemma 2.2.1. *For $a, c > 0$ and $\alpha \geq 0$,*

$$\int_0^\infty s^{\alpha-2}\exp\left\{-\frac{a}{s} - cs\right\}\mathrm{d}s = 2\left(\frac{a}{c}\right)^{\frac{\alpha}{2}-\frac{1}{2}}\mathrm{K}_{\alpha+1}\left(2\sqrt{ac}\right) - \frac{2\alpha}{c}\left(\frac{a}{c}\right)^{\frac{\alpha}{2}-1}\mathrm{K}_\alpha\left(2\sqrt{ac}\right).$$

Remark 2.2.1. *Lemma 2.2.1 and identity (10.2.17) from Abramowitz and Stegun (1964, p. 444), give that*

$$\int_0^\infty s^{-\frac{1}{2}}\exp\left\{-\frac{a}{s} - cs\right\}\mathrm{d}s = \left(\frac{\pi}{c}\right)^{\frac{1}{2}}\mathrm{e}^{-2\sqrt{ac}},$$

$$\int_0^\infty s^{-\frac{3}{2}}\exp\left\{-\frac{a}{s} - cs\right\}\mathrm{d}s = \left(\frac{\pi}{a}\right)^{\frac{1}{2}}\mathrm{e}^{-2\sqrt{ac}},$$

$$\textit{and}\quad \int_0^\infty s^{-\frac{5}{2}}\exp\left\{-\frac{a}{s} - cs\right\}\mathrm{d}s = \left(\frac{c}{a} + \frac{1}{2a}\left(\frac{c}{a}\right)^{\frac{1}{2}}\right)\left(\frac{\pi}{c}\right)^{\frac{1}{2}}\mathrm{e}^{-2\sqrt{ac}}.$$

Proof.

$$\begin{aligned}
&\int_0^\infty s^{\alpha-2}\exp\left\{-\frac{a}{s} - cs\right\}\mathrm{d}s \\
&= \frac{c}{a}\int_0^\infty s^{\alpha}\exp\left\{-\frac{a}{s} - cs\right\}\mathrm{d}s \\
&\quad - \frac{s^\alpha}{a}\exp\left\{-\frac{a}{s} - cs\right\}\Big|_0^\infty + \int_0^\infty\left(s^{\alpha-2} - \frac{c}{a}s^\alpha\right)\exp\left\{-\frac{a}{s} - cs\right\}\mathrm{d}s \\
&= \frac{c}{a}\int_0^\infty s^{\alpha}\exp\left\{-\frac{a}{s} - cs\right\}\mathrm{d}s - \frac{\alpha}{a}\int_0^\infty s^{\alpha-1}\exp\left\{-\frac{a}{s} - cs\right\}\mathrm{d}s \\
&= 2\left(\frac{a}{c}\right)^{\frac{\alpha}{2}-\frac{1}{2}}\int_0^\infty \exp\left\{2\sqrt{ac}\cosh x\right\}\cosh(\alpha+1)x\,\mathrm{d}x \\
&\quad - \frac{2\alpha}{c}\left(\frac{a}{c}\right)^{\frac{\alpha}{2}-1}\int_0^\infty \exp\left\{2\sqrt{ac}\cosh x\right\}\cosh\alpha x\,\mathrm{d}x
\end{aligned}$$

$$= 2\left(\frac{a}{c}\right)^{\frac{\alpha}{2}-\frac{1}{2}} \mathrm{K}_{\alpha+1}\left(2\sqrt{ac}\right) - \frac{2\alpha}{c}\left(\frac{a}{c}\right)^{\frac{\alpha}{2}-1} \mathrm{K}_{\alpha}\left(2\sqrt{ac}\right),$$

where we have used identity (9.6.24) from Abramowitz and Stegun (1964, p. 376) and the substitution $s = (a/c)^{\frac{1}{2}}\mathrm{e}^{x}$. □

Proof of Theorem 2.2.1. By the law of total probability,

$$f_t(x) = \int_0^\infty \frac{1}{2\pi \det(\Sigma)^{\frac{1}{2}} s} \exp\left\{-\frac{\langle x - s\mu, \Sigma^{-1}(x - s\mu)\rangle}{2s}\right\} \frac{b^{bt} s^{bt-1}}{\Gamma(bt)} \mathrm{e}^{-bs}\, \mathrm{d}s.$$

By Theorem 4.2 of Cont and Tankov (2004),

$$\Pi(\mathcal{B}) = \int_0^\infty \int_{\mathcal{B}} \frac{1}{2\pi \det(\Sigma)^{\frac{1}{2}} s} \exp\left\{-\frac{\langle x - s\mu, \Sigma^{-1}(x - s\mu)\rangle}{2s}\right\} \mathrm{d}x \frac{b\mathrm{e}^{-bs}}{s}\, \mathrm{d}s,$$

where $\mathcal{B}$ is a Borel subset of $\mathbb{R}^2 \setminus \{0\}$. The result follows from two applications of Lemma 2.2.1. □

The $\alpha\beta$-Variance Gamma process of Semeraro (2008) takes a slightly different approach, and uses a multivariate subordinator to subordinate Gaussian processes with independent components to give a different multidimensional VG process. Multivariate subordinators are a relatively recent development. Barndoff-Nielsen, Pedersen, and Sato (2001) gives the Lévy triplets for subordinated processes where a multivariate subordinator is used with a subordinant with independent components. For the $\alpha\beta - VG$ process, take $a, b, \alpha, \beta > 0$ satisfying $0 < \alpha < \frac{b}{a}$ and $0 < \beta < \frac{b}{a}$ and three independent Gamma processes $X_1(t)$, $X_2(t)$, and $Z(t)$, $t \geq 0$, such that

$$X_1(1) \sim \Gamma\left(\frac{b}{\alpha} - a, \frac{b}{\alpha}\right),\ X_2(1) \sim \Gamma\left(\frac{b}{\beta} - a, \frac{b}{\beta}\right),\ \text{and } Z(1) \sim \Gamma(a, b).$$

Semeraro (2008) shows that these can be used to define a multivariate subordinator

made up of dependent Gamma processes, as follows:

$$G(t) = (G_1(t), G_2(t)) = (X_1(t) + \alpha Z(t), X_2(t) + \beta Z(t)) .$$

Define a Brownian process $B(t) = (B_1(t), B_2(t))$ with mean vector $\mu = (\mu_1, \mu_2) \in \mathbb{R}^2$ and covariance matrix $\Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$. The $\alpha\beta$-Variance Gamma process is then given by

$$L(t) = (B_1(G_1(t)), B_2(G_2(t))) . \tag{2.2.3}$$

The next theorem gives its Lévy measure, which Semeraro (2008) does not provide explicitly.

Theorem 2.2.2. *The Lévy measure of the process defined in equation* (2.2.3) *is given by*

$$\begin{aligned} &\Pi(\mathcal{B}) \qquad (2.2.4)\\ &= \frac{a}{\pi}\left(\frac{\alpha\mu_1^2}{\sigma_1^2} + \frac{\beta\mu_2^2}{\sigma_2^2} + 2b\right)^{\frac{1}{2}} \int_{\mathcal{B}} \left(\beta\sigma_2^2 x_1^2 + \alpha\sigma_1^2 x_2^2\right)^{-\frac{1}{2}} \exp\left\{\frac{\mu_1 x_1}{\sigma_1^2} + \frac{\mu_2 x_2}{\sigma_2^2}\right\} \\ &\quad \times \mathrm{K}_1\left(\left(\frac{\alpha\mu_1^2}{\sigma_1^2} + \frac{\beta\mu_2^2}{\sigma_2^2} + 2b\right)^{\frac{1}{2}} \left(\frac{x_1^2}{\alpha\sigma_1^2} + \frac{x_2^2}{\beta\sigma_2^2}\right)^{\frac{1}{2}}\right) \mathrm{d}(x_1, x_2) \\ &\quad + \left(\frac{b}{\alpha} - a\right) \int_{\mathcal{B}_1} \frac{1}{|x|} \exp\left\{\frac{\mu_1 x}{\sigma_1^2} - \left(\frac{\mu_1^2}{\sigma_1^2} + 2\frac{b}{\alpha}\right)^{\frac{1}{2}} \frac{|x|}{\sigma_1}\right\} \mathrm{d}x \\ &\quad + \left(\frac{b}{\beta} - a\right) \int_{\mathcal{B}_2} \frac{1}{|x|} \exp\left\{\frac{\mu_2 x}{\sigma_2^2} - \left(\frac{\mu_2^2}{\sigma_2^2} + 2\frac{b}{\beta}\right)^{\frac{1}{2}} \frac{|x|}{\sigma_2}\right\} \mathrm{d}x, \end{aligned}$$

where $\mathcal{B}_1 = \{b_1 : (b_1, 0) \in \mathcal{B}\}$, $\mathcal{B}_2 = \{b_2 : (0, b_2) \in \mathcal{B}\}$, *and* $\mathcal{B}$ *is a Borel subset of* $\mathbb{R}^2 \setminus \{0\}$.

Proof. Define $s = (s_1, s_2) \in \mathbb{R}^{+2}$ and $\rho_s(\mathcal{B}) = \mathrm{P}((B_1(s_1), B_2(s_2)) \in \mathcal{B})$ for Borel

$\mathcal{B} \subset \mathbb{R}^2 \setminus \{0\}$. The Lévy measure of $L(t)$ is derived in Semeraro (2006) to be

$$\Pi(\mathcal{B}) = \int_{\mathbb{R}^{+2}} \rho_s(\mathcal{B}) \Pi_G(\mathrm{d}s). \tag{2.2.5}$$

Define the Lévy measures Π_G, Π_{X_1}, Π_{X_2}, and Π_Z for the processes $G(t)$, $X_1(t)$, $X_2(t)$, and $Z(t)$ respectively. We can evaluate Π_G as follows. For a Borel set $\mathcal{A} \subset \mathbb{R}^{+2}$,

$$\Pi_G(\mathcal{A}) = \mathrm{E} \sum_{0<s\leq 1} \mathbf{1}\left((\Delta G_1(s), \Delta G_2(s)) \in \mathcal{A}\right).$$

As $X_1(t)$, $X_2(t)$, and $Z(t)$ are independent jump processes, they jump simultaneously with probability zero. So, for $g_1 \geq 0$ and $g_2 \geq 0$,

$$\begin{aligned}
&\Pi_G((g_1, \infty) \times (g_2, \infty)) \\
&= \mathrm{E} \sum_{0<s\leq 1} (\Delta X_1(s) + \alpha \Delta Z(s) > g_1, \Delta X_2(s) + \beta \Delta Z(s) > g_2) \\
&= \mathrm{E} \sum_{0<s\leq 1} [\mathbf{1}(\alpha \Delta Z(s) > g_1, \beta \Delta Z(s) > g_2, \Delta Z > 0) \\
&\quad + \mathbf{1}(\Delta X_1(s) > g_1, \Delta X_2(s) = g_2, \Delta Z(s) = 0) \\
&\quad + \mathbf{1}(\Delta X_1(s) = g_1, \Delta X_2(s) > g_2, \Delta Z(s) = 0)] \\
&= \Pi_Z\left(\left(\frac{g_1}{\alpha} \wedge \frac{g_2}{\beta}, \infty\right)\right) + \Pi_{X_1}((g_1, \infty))\delta_0(\{g_2\}) + \delta_0\left(\{g_1\}\right) \Pi_{X_2}((g_2, \infty)).
\end{aligned}$$

This can be used to rewrite (2.2.5) as

$$\begin{aligned}
&\Pi(\mathcal{B}) \\
&= \int_{\mathbb{R}^+} \rho_{(\alpha s, \beta s)}(\mathcal{B}) \Pi_Z(\mathrm{d}s) + \int_{\mathbb{R}^+} \rho_{(s,0)}(\mathcal{B}) \Pi_{X_1}(\mathrm{d}s) + \int_{\mathbb{R}^+} \rho_{(0,s)}(\mathcal{B}) \Pi_{X_2}(\mathrm{d}s)
\end{aligned}$$

$$= \frac{a}{2\pi \det(A\Sigma)^{\frac{1}{2}}} \int_{\mathbb{R}^+} \int_{\mathcal{B}} \frac{1}{s} \exp\left\{ -\frac{\langle x - sA\mu, (A\Sigma)^{-1}(x - sA\mu)\rangle}{2s} \right\} \mathrm{d}x \frac{\mathrm{e}^{-bs}}{s} \mathrm{d}s$$
$$+ \frac{\frac{b}{\alpha} - a}{(2\pi)^{\frac{1}{2}}\sigma_1} \int_{\mathbb{R}^+} \int_{\mathcal{B}_1} s^{-\frac{1}{2}} \exp\left\{ -\frac{(x - s\mu_1)^2}{2s\sigma_1^2} \right\} \mathrm{d}x \frac{\mathrm{e}^{-\frac{b}{\alpha}s}}{s} \mathrm{d}s$$
$$+ \frac{\frac{b}{\beta} - a}{(2\pi)^{\frac{1}{2}}\sigma_2} \int_{\mathbb{R}^+} \int_{\mathcal{B}_2} s^{-\frac{1}{2}} \exp\left\{ -\frac{(x - s\mu_2)^2}{2s\sigma_2^2} \right\} \mathrm{d}x \frac{\mathrm{e}^{-\frac{b}{\beta}s}}{s} \mathrm{d}s,$$

where $A = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$, $\mathcal{B}_1 = \{a_1 : (a_1, 0) \in \mathcal{B}\}$, and $\mathcal{B}_2 = \{a_2 : (0, a_2) \in \mathcal{B}\}$. The result now follows from three applications of Lemma 2.2.1. □

It is possible to calculate the transition probability density function and Lévy measure for a bivariate NIG process generated by subordinating a general bivariate Gaussian process with a common Inverse Gaussian process. Take a two dimensional Gaussian process $B(t)$ with drift vector μ and covariance matrix Σ and an Inverse Gamma process $I(t)$ such that $I(1) \sim \mathrm{IG}\left(\frac{b}{\sqrt{2\pi}}, \frac{b^2}{2}\right)$. The following theorem gives the transition probability density function, which appears in Barndorff-Nielsen (1997) with a different parameterisation, and the Lévy measure of $L(t) = B(I(t))$, which does not seem to be present in the literature.

Theorem 2.2.3. *The transition probability density function and Lévy measure for $L(t) = B(I(t))$, as defined above, are*

$$f_t(x) = \frac{bt}{2\pi \det(\Sigma)^{\frac{1}{2}}} \left[\frac{\left(\langle \mu, \Sigma^{-1}\mu\rangle + b^2\right)^{\frac{1}{2}}}{\langle x, \Sigma^{-1}x\rangle + b^2t^2} + \frac{1}{\left(\langle x, \Sigma^{-1}x\rangle + b^2t^2\right)^{\frac{3}{2}}} \right]$$
$$\times \exp\left\{ \langle \mu, \Sigma^{-1}x\rangle + b^2t - 2\left(\langle \mu, \Sigma^{-1}\mu\rangle + b^2\right)^{\frac{1}{2}} \left(\langle x, \Sigma^{-1}x\rangle + b^2t^2\right)^{\frac{1}{2}} \right\}$$

and

$$\Pi(\mathrm{d}x) = \frac{b}{2\pi \det(\Sigma)^{\frac{1}{2}}} \left[\frac{\left(\langle \mu, \Sigma^{-1}\mu\rangle + b^2\right)^{\frac{1}{2}}}{\langle x, \Sigma^{-1}x\rangle} + \frac{1}{\langle x, \Sigma^{-1}x\rangle^{\frac{3}{2}}} \right]$$
$$\times \exp\left\{ \langle \mu, \Sigma^{-1}x\rangle - 2\left(\langle \mu, \Sigma^{-1}\mu\rangle + b^2\right)^{\frac{1}{2}} \langle x, \Sigma^{-1}x\rangle^{\frac{1}{2}} \right\} \mathrm{d}x.$$

Proof. The proof proceeds in exactly the same fashion as for Theorem 2.2.1. □

As for one-dimensional Lévy processes, multidimensional Lévy processes can be specified by characterising their Lévy triplets. When modelling multidimensional random variables, the dependence structure of the variables can be separated from their marginal distributions through the use of a *copula* function. For instance, for a bivariate random variable with probability measure ν, cumulative distribution function $F(x_1, x_2) = \nu((-\infty, x_1) \times (-\infty, x_2))$, and marginal cumulative distribution functions $F_1(x) = F(x, \infty)$ and $F_2(x) = F(\infty, x)$, a copula $C : [0,1]^2 \mapsto [0,1]$ can be defined by $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$. Lévy copulas can be used in an analogous way to describe multidimensional Lévy measures in terms of one-dimensional Lévy measures, with the added complication that a Lévy measure is not necessarily bounded. For a detailed development of Lévy copulas see Chapter 5 of Cont and Tankov (2004). We give one illustrative example.

Example 5.9 in Cont and Tankov (2004) describes a multivariate VG model by defining its Lévy measure Π in the following fashion. Take two one-dimensional Lévy measures Π_1 and Π_2 of the form given in (2.2.1). Define the marginal tail integrals for $j \in \{1, 2\}$: $U_j^+(x) = \Pi_j([x, \infty))$, $x > 0$, and $U_j^-(x) = -\Pi_j((-\infty, x])$, $x < 0$, with $U^+(\infty) = U^-(\infty) = 0$ and $U^+(0) = -U^-(0) = \infty$. Also define the tail integrals of Π:

$$
\begin{aligned}
U^{++}(x_1, x_2) &= \Pi([x_1, \infty) \times [x_1, \infty)), && x_1 > 0, x_2 > 0, \\
U^{+-}(x_1, x_2) &= -\Pi([x_1, \infty) \times (-\infty, x_2]), && x_1 > 0, x_2 < 0, \\
U^{-+}(x_1, x_2) &= -\Pi((-\infty, x_1] \times [x_2, \infty)), && x_1 < 0, x_2 > 0, \\
\text{and} \quad U^{--}(x_1, x_2) &= \Pi((-\infty, x_1] \times (-\infty, x_2), && x_1 < 0, x_2 < 0,
\end{aligned}
$$

with $U^{\cdot\cdot}(\pm\infty, \cdot) = U^{\cdot\cdot}(\cdot, \pm\infty) = 0$, and

$$\begin{aligned} U^{++}(x,0) - U^{+-}(x,0) &= U_1^+(x), \\ U^{-+}(x,0) - U^{--}(x,0) &= U_1^-(x), \\ U^{++}(0,x) - U^{-+}(0,x) &= U_2^+(x), \\ \text{and} \quad U^{+-}(0,x) - U^{--}(0,x) &= U_2^-(x). \end{aligned}$$

The tail integrals are now linked to the marginal tail integrals via the Lévy copula $F : [-\infty, \infty]^2 \mapsto [-\infty, \infty]$:

$$\begin{aligned} U^{++}(x_1,x_2) &= F(U_1^+(x_1), U_2^+(x_2)), && x_1 \geq 0, x_2 \geq 0, \\ U^{+-}(x_1,x_2) &= F(U_1^+(x_1), U_2^-(x_2)), && x_1 \geq 0, x_2 \leq 0, \\ U^{-+}(x_1,x_2) &= F(U_1^-(x_1), U_2^+(x_2)), && x_1 \leq 0, x_2 \geq 0, \\ \text{and} \quad U^{--}(x_1,x_2) &= F(U_1^-(x_1), U_2^-(x_2)), && x_1 \leq 0, x_2 \leq 0. \end{aligned}$$

Finally, Π is specified by giving F the following parametric form:

$$F(u,v) = \begin{cases} \left(u^{-\phi^+} + v^{-\phi^+}\right)^{-\frac{1}{\phi^+}}, & u > 0, v > 0, \\ \left(|u|^{-\phi^-} + |v|^{-\phi^-}\right)^{-\frac{1}{\phi^-}}, & u < 0, v < 0, \\ 0, & uv \leq 0, \end{cases}$$

for some $\phi^+, \phi^- > 0$.

An approach that does not seem to have been taken in the literature, but is suggested here, is to define a multidimensional Lévy process by stipulating its Lévy measure directly. Candidates for this method might be drawn, for instance, from the class of Lamperti stable processes. Lamperti stable processes in one dimension have been suggested for the purpose of financial modelling in Sengul (2008), where

the Lévy measure is defined as

$$\Pi(\mathrm{d}x) = \left[\frac{c_+ \mathrm{e}^{\beta x}}{(\mathrm{e}^x - 1)^{\alpha+1}} \mathbf{1}_{x>0} + \frac{c_- \mathrm{e}^{\rho x}}{(1 - \mathrm{e}^x)^{\alpha+1}} \mathbf{1}_{x<0}\right] \mathrm{d}x, \quad \alpha \in (0, 2).$$

The Lévy measure for a Lamperti stable process in d dimensions is given in Caballero, Pardo, and Pérez (2008) as

$$\Pi(B) = \int_{S^{d-1}} \sigma(\mathrm{d}\xi) \int_0^\infty \mathbf{1}_B(r\xi) \mathrm{e}^{rf(\xi)} (\mathrm{e}^r - 1)^{-(\alpha+1)} \mathrm{d}r$$

for B a Borel set in $\mathbb{R}^d$, where $\alpha \in (0, 2)$, S^{d-1} is the unit sphere in $\mathbb{R}^d$, σ is a non-zero finite measure on S^{d-1}, and $f : S^{d-1} \mapsto \mathbb{R}$ is a measurable function such that $\sup_{\xi \in S^{d-1}} f(\xi) < \alpha + 1$. Modelling with these processes would involve parametrising σ and f. This is left for future investigation.

2.3 Calculating the Price of American Options Using Lévy Models

This section provides a brief overview of the different methods that have been used to numerically evaluate $v(t, S)$, the solution to (2.1.3), and have been reported in the literature. Most are based on the numerical solution of a partial integro-differential equation (PIDE) that $v(t, S)$ can be shown to verify. Other techniques include lattice methods, which are similar to the PIDE approach but generally have a more probabilistic interpretation, and the ubiquitous Monte-Carlo approach. Cont and Tankov (2004, Chapter 12) provides a comprehensive introduction to PIDEs as applied to pricing with Lévy models, and many references. For a slightly more recent review of PIDEs and Monte Carlo methods see Schoutens (2006).

The PIDE in question can be presented in several equivalent forms. For example, take an American put on a single asset. That is, assume that (2.1.1) holds with $d = 1$

and that $h(t,S) = (K-S)^+$. We can say that $v(t,S)$ verifies

$$\left(-\frac{\partial v}{\partial t}(t,S) - \mathscr{L}v(t,S) + rS\right) \wedge \left(v(t,S) - (K-S)^+\right) = 0 \tag{2.3.1}$$

$$\text{under the boundary condition} \quad v(T,S) = (K-S)^+, \tag{2.3.2}$$

where

$$\begin{aligned}\mathscr{L}v(t,S) = rS\frac{\partial v}{\partial S}(t,S) + \frac{\sigma^2 S^2}{2}\frac{\partial^2 v}{\partial S^2}(t,S) \\ + \int_{\mathbb{R}} \left[v\left(t, S\mathrm{e}^y\right) - v(t,S) + S\left(\mathrm{e}^y - 1\right)\frac{\partial v}{\partial S}(t,S)\right] \Pi(\mathrm{d}y)\end{aligned}$$

for $(t,S) \in [0,T] \times \mathbb{R}^+$ (Cont and Tankov 2004, Section 12.1.3). $\mathscr{L}$ is known as the *infinitesimal generator*.

Important concepts for American options are the *exercise region* $\mathcal{E} \subset [0,T] \times \mathbb{R}^+$, in which the option should be exercised immediately, and its complement the *continuation region* $\mathcal{C}$, in which an owner should hold the option. For an American put on one asset it can be shown (Cont and Tankov 2004, Section 12.1.3) that these regions have a particularly simple representation and that there exists a function known as the *exercise boundary* $b : [0,T] \mapsto \mathbb{R}^+$ such that

$$\mathcal{E} = \left\{(t,S) : v(t,S) = (K-S)^+\right\} = \left\{(t,S) : S \leq b(t)\right\} \quad \text{and}$$

$$\mathcal{C} = \left\{(t,S) : v(t,S) > (K-S)^+\right\} = \left\{(t,S) : S > b(t)\right\}.$$

If $v(t,S)$ is continuously differentiable with respect to S on the exercise boundary, a condition known as *smooth pasting*, then $v(t,S)$ can be solved using finite difference methods as a free boundary problem. An example of this technique for an American put with a jump-diffusion model can be found in Pham (1997). Note that finite difference methods for PIDEs are more complicated than those for PDEs; the integral term in the infinitesimal generator is a nonlocal term, that is that for a particular

(t, S), $v(t, S)$ is related to $v(t, \cdot)$ over the positive real line, not just to local derivatives of v. If the Lévy model under consideration is a pure jump process or has infinite activity, then smooth pasting may fail (Cont and Tankov 2004, Remark 12.3) and other methods must be considered.

Another way to address the problem is to phrase it as a so-called linear complementarity problem. This method is used for example in Almendral (2005) and Almendral and Oosterlee (2007) to price an American put assuming a restricted CGMY model and a Variance Gamma model respectively. In that paper the problem is stated for $\tau = T - t$, $x = \ln S$, and $u(\tau, x) = v(t, S)$ as:

$$\begin{aligned}
\frac{\partial u}{\partial \tau}(\tau, x) - \mathscr{L}_{\log} u(\tau, x) &= 0, && (\tau, x) \in (0, T) \times \mathbb{R}, \\
u(\tau, x) &\geq \psi(x), && (\tau, x) \in [0, T] \times \mathbb{R}, \\
\left(\frac{\partial u}{\partial \tau}(\tau, x) - \mathscr{L}_{\log} u(\tau, x) \right) (u(\tau, x) - \psi(x)) &= 0, && (\tau, x) \in (0, T) \times \mathbb{R}, \\
u(0, x) &= \psi(x), && x \in \mathbb{R},
\end{aligned}$$

where

$$\begin{aligned}
\mathscr{L}_{\log} u(\tau, x) = \frac{\sigma^2}{2} \frac{\partial^2 u}{\partial x^2}(\tau, x) + \left(r - \frac{\sigma^2}{2} \right) \frac{\partial u}{\partial x}(\tau, x) - r u(\tau, x) \\
+ \int_{\mathbb{R}} \left[u(\tau, x + y) - u(\tau, x) + (e^y - 1) \frac{\partial u}{\partial x}(\tau, x) \right] \Pi(\mathrm{d}y),
\end{aligned}$$

and $\psi(x) = (K - e^x)^+$. If the time interval $[0, T]$ is discretised as $\{0, \Delta\tau, \ldots, N\Delta\tau\}$, $N \in \mathbb{N}$, and the partial derivative of u with respect to τ is replaced with its finite difference approximation,

$$\frac{\partial u}{\partial \tau}(j\Delta\tau, x) \approx \frac{u((j+1)\Delta\tau, x) - u(j\Delta\tau, x)}{\Delta\tau}, \quad j \in \{0, \ldots, N-1\},$$

then we have a sequence of linear complementarity problems that allows $u((j+1)\Delta\tau, \cdot)$ to be determined from $u(j\Delta\tau, \cdot)$ iteratively from $u(0, x) = \psi(x)$

to $u(N\Delta\tau, x)$. In the case of Almendral (2005), each problem in the sequence is solved using another finite difference scheme. As the CGMY model possesses infinite activity, the integral term in the PIDE includes an integrable singularity near zero that makes discretisation of the integral difficult. This is overcome in Almendral (2005) by using instead a model that approximates the CGMY model by truncating the Lévy measure away from zero and compensating by adding a small diffusion component. For details on this common technique see Cont and Tankov (2004, Section 6.4). Almendral and Oosterlee (2007) solve the PIDE using the same finite difference approach, but deal with the singularity by simply truncating the Lévy measure away from zero in such a way that the error so introduced disappears asymptotically.

A number of numerical techniques that have been applied to solving the PIDE can be categorised as Galerkin methods. See Cont and Tankov (2004, Section 12.6) for an introduction to pricing American options in this fashion. The key difference between Galerkin methods and the finite difference method just described lies in the way in which the (log) price variable is discretised. Rather than approximating the price function on a regular grid on $[0, T] \times \mathbb{R}$, Galerkin methods express the approximation using the first $n \in \mathbb{N}$ terms of the expansion $u(\tau, x) = \sum_{i \geq 1} a_i(\tau) e_i(x)$, where $\{e_i\}_{i \in \mathbb{N}}$ is a complete basis of an appropriate Hilbert space. These expansions can be found using what is known as a variational formulation of the problem.

A few different choices of bases are present in the literature. Matache, Nitsche, and Schwab (2005) and Matache, Schwab, and Wihler (2005) use a wavelet basis that allows the driving Lévy process to have infinite activity and makes use of the resulting sparse matrices to speed calculation. Matache et al. (2005) prices Americans with some technical restrictions on the Lévy measure and the option payoff that nonetheless allow a restricted CGMY model and payoffs that grow polynomially to infinity and are not necessarily continuous.

The Analytic Method of Lines that is employed by several authors (Cont and

Tankov 2004, Section 12.5) can be viewed as a Galerkin method with a discrete Fourier basis (Cont and Tankov 2004, Section 12.6.2), although this method uses a technique called Wiener-Hopf factorisation instead of a variational formation to calculate the approximations to $u(\tau, x)$. The method of lines is used, for example, in Levendorskiĭ (2004) to price American puts using Lévy models that are again subject to some technical restrictions, this time to allow Wiener-Hopf factorisation to be used.

While all of the PIDE methods described so far use a backward Euler scheme to propagate prices from the payoff at maturity back to the starting time, it doesn't have to be done that way. Carr and Hirsa (2003) price American puts using a forward PIDE with strike and maturity as independent variables. They also assume a smooth pasting condition and give an example where the Lévy model is a Variance Gamma model with an added diffusion component.

Finite difference methods can be classified as either explicit or implicit. All of the methods mentioned above are explicit, because for some time partition $\{t_0 = 0, t_1, \ldots, t_N = T\}$ they recursively calculate each numerical solution to $v(t_k, \cdot)$ using the solution to $v(t_{k+1}, \cdot)$ or vice versa. Implicit methods function by solving a set of discrete equations that relate the value function at each grid node to that at other nodes in the grid. In general, implicit methods exhibit more desirable numerical convergence properties than explicit methods.

Implicit methods are most commonly reduced to matrix inversion problems, typically of large, sparse matrices, but it is difficult to formulate a suitable matrix equation under the constraint imposed by American options that $v(t, x) \geq h(t, x)$, which goes some way to explaining why all of the above methods are explicit. However, the implicit method is used to price Americans using jump-diffusion processes in d'Halluin, Forsyth, and Labahn (2004). This paper uses an iterative approach to implement the penalty method. That is, techniques exist to find iteratively better

solutions to the nonlinear PIDE,

$$\frac{\partial v}{\partial t}(t,S) = -\mathscr{L}v(t,S) + rS + \rho\left((K-S)^+ - v(t,S)\right)^+,$$

which also solves (2.3.2) and, under the limit $\rho \to \infty$, guarantees that $v(t,x) \geq h(t,x)$. In practice the penalty coefficient ρ is chosen to be a large positive constant. In d'Halluin et al. (2004), an American put and American butterfly are presented as examples, convergence of the numerical scheme is proved under certain criteria, and empirically established quadratic convergence rates are reported.

Lattice methods can be categorised as explicit finite difference schemes (Cont and Tankov 2004, Section 12.3.2), but represent a more probabilistic approach to the pricing problem than those listed above. For an explicit example of the relationship between lattice and finite difference methods for Gaussian processes see Hull (2003, p. 422). Lattice methods are so called because they assume a simplified market model in which the log prices of risky assets reside on a discrete grid and on which the price of American and exotic options can be calculated explicitly using dynamic programming algorithms. As explicit schemes, if implemented naïvely, they converge more slowly than their PIDE counterparts, but offer financial transparency by maintaining arbitrage relationships amongst instruments (Cont and Tankov 2004, Section 12.3.2) and, if properly formulated, offer very broad applicability (Maller, Solomon, and Szimayer 2006).

Këllezi and Webber (2004) introduced several lattice models with the following common structure. Assume the definition of $L(t)$ and $v(t,S)$ given in Section 2.1 in one dimension. Take $\Delta t, \Delta l \in \mathbb{R}^+$ and $N, U, D, D^{\text{MAX}}, U^{\text{MAX}} \in \mathbb{N}$ with $N\Delta t = T$, $\mathcal{N} = \{0,\ldots,N-1\}$, $\mathcal{J} = \left\{-D^{\text{MAX}},\ldots,U^{\text{MAX}}\right\}$, $\mathcal{K} \subset \mathcal{J}$, $\min\mathcal{K} = -D$, $\max\mathcal{K} = U$, and $0 \in \mathcal{K}$. Define the discrete stochastic process $\widetilde{L}(i\Delta t) \in \{j\Delta l\}_{j\in\mathcal{J}}$, $i \in \mathcal{N} \cup \{N\}$

and $L(0) = 0$, and its transition probability mass function

$$p_{k-j} = \mathrm{P}\left(\widetilde{L}((i+1)\Delta t) = k\Delta l \middle| \widetilde{L}(i\Delta t) = j\Delta l\right), \quad i \in \mathcal{N}, (k-j) \in \mathcal{K}, k \in \mathcal{J}.$$

The values for $\widetilde{L}(t) \in \left\{j\Delta l : j \in \mathcal{J} \setminus \left\{-D^{\mathrm{MAX}} + D, U^{\mathrm{MAX}} - U, \right\}\right\}$ are not covered by the above definition, as p_k or $\widetilde{L}(t)$ must be modified in this region to ensure the truncated range of $\widetilde{L}(t)$. The above construction is formulated in such a way that $\widetilde{L}(t)$ converges in some sense to $L(t)$ as $(\Delta t, \Delta l) \to 0$ for $\widetilde{L}(t) \in \left\{\left(-D^{\mathrm{MAX}} + D\right)\Delta l, \ldots, \left(U^{\mathrm{MAX}} - U\right)\Delta l\right\}$. The risky asset price is then modelled as

$$\widetilde{S}(i\Delta t) = S(0)\exp\left\{\widetilde{\omega} i\Delta t + \widetilde{L}(i\Delta t)\right\},$$

where $S(0) \in \mathbb{R}^+$ and the compensator $\widetilde{\omega}$ is defined so that $\mathrm{E}\exp\left\{\widetilde{\omega}\Delta t + \widetilde{L}(\Delta t)\right\} = \mathrm{e}^{r\Delta t}$. Note that this parameterisation of $\widetilde{\omega}$ differs to that in Këllezi and Webber (2004) by r, but brings the current definition in line with our setup. For a payoff function $h : [0,t] \times \mathbb{R}^+ \mapsto \mathbb{R}^+$, the recursion relation for $i \in \mathcal{N}$,

$$\widetilde{v}(i\Delta t, s_i) = \max\left(h(i\Delta t, s_i), \mathrm{e}^{-r\Delta t}\sum_{k\in\mathcal{K}} p_k \widetilde{v}\left((i+1)\Delta t, s_i \mathrm{e}^{k\Delta l}\right)\right), \quad (2.3.3)$$

with $\widetilde{v}(N\Delta t, s_N) = h(N\Delta t, s_N)$,

where $s_i \in \left\{\exp\left\{\widetilde{\omega} i\Delta t + j\Delta l\right\}\right\}_{j\in\mathcal{J}}$, solves the pricing equation

$$\widetilde{v}(i\Delta t, s_i) = \sup_{\tau \in \mathcal{S}^{\widetilde{L}}_{i\Delta t, T}} \mathrm{E}\left(\mathrm{e}^{-r(\tau - i\Delta t)} h\left(\tau, s_i \exp\left\{\widetilde{\omega}\tau + \widetilde{L}(\tau)\right\}\right)\right), \quad (2.3.4)$$

where $\mathcal{S}^{\widetilde{L}}_{i\Delta t, T}$ is the set of stopping times in $\{i\Delta t, \ldots, N\Delta t\}$ for the natural filtration of $\widetilde{L}(t)$. This framework can be used to define several lattice models by specifying different functions for p_k. The main example given in Këllezi and Webber (2004)

sets

$$\begin{aligned} p_k &= f_{\Delta t}(k\Delta l)n_k, \quad k \in \mathcal{K} \setminus \{K\} \quad \text{and} \\ p_K &= 1 - \sum_{k \in K \setminus \{K\}} p_k, \end{aligned} \tag{2.3.5}$$

where $K = \arg\max_{k \in \mathcal{K}} f_{\Delta t}(k\Delta l)$, f_t is the transition probability density function for $L(t)$, and n_k are normalising coefficients defined as follows. Order the elements of

$$\mathcal{K} = \{k_{-d} = -D, k_{-d+1}, \ldots, k_0 = 0, \ldots, k_{u-1}, k_u = U\}, \quad u, d \in \mathbb{N}.$$

n_k is now defined as

$$\begin{aligned} n_{k_q} &= \frac{1}{2}(k_{q+1} - k_{q-1})\Delta l, \quad q \in \{-d+1, \ldots, u-1\} \\ n_{k_u} &= \frac{1}{2}(k_u - k_{u-1})\Delta l, \quad \text{and} \\ n_{k_{-d}} &= \frac{1}{2}(k_{-d+1} - k_{-d})\Delta l. \end{aligned}$$

Thus p_k is an approximation to $f_{\Delta t}$. Këllezi and Webber define several other models by deriving p_k from the Lévy measure of $L(t)$ for finite activity processes, using the law of total probability for processes that admit a subordinated Brownian representation, and using *time copulas*, which offer an alternative method for describing the time evolution of a transition probability density function.

The above model is only useful if the price $\widetilde{v}(i\Delta t, \widetilde{S}(i\Delta t))$ of (2.3.4) converges in some sense to the price $v(t, S(t))$ of (2.1.3). Këllezi and Webber empirically demonstrate convergence for Bermudan options, that is options that can be exercised at discrete intervals, for the Variance Gamma and Normal Inverse Gaussian models, but are unable to demonstrate convergence for Americans. Further, it has not been shown theoretically that $\widetilde{v}(i\Delta t, \widetilde{S}(i\Delta t))$ necessarily does converge to $v(t, S(t))$ for any of the models suggested in Këllezi and Webber (2004).

Remark 2.3.1. *In Section 3.5, it will be shown that a model very similar to* (2.3.5) *does converge theoretically for all Lévy processes and some American options. It will now be explained how* (2.3.5) *can be modified to make it consistent with the model in Section 3.5, and the formal convergence result given there will be stated.*

Start with the simplifying assumption that $\mathcal{K} = \{k_{-d} = -d, -d+1, \ldots, u-1, k_u = u\}$ *and define a new model as*

$$\begin{aligned} p_k &= \int_{(k-\frac{1}{2})\Delta l}^{(k+\frac{1}{2})\Delta l} f_{\Delta t}(x)\,\mathrm{d}x, \\ p_u &= \int_{(u-\frac{1}{2})\Delta l}^{\infty} f_{\Delta t}(x)\,\mathrm{d}x, \quad \textit{and} \\ p_{-d} &= \int_{-\infty}^{(-d+\frac{1}{2})\Delta l} f_{\Delta t}(x)\,\mathrm{d}x. \end{aligned}$$

To see the similarity between this definition and (2.3.5), *note that for* $k \in \{-d+1, \ldots, u-1\}$,

$$f_{\Delta t}(k\Delta l)n_k = f_{\Delta t}(k\Delta l)\Delta l = \int_{(k-\frac{1}{2})\Delta l}^{(k+\frac{1}{2})\Delta l} f_{\Delta t}(x)\,\mathrm{d}x - \frac{\Delta l^2}{4} f'_{\Delta t}(\xi),$$

for some $\xi \in \left[\left(k-\frac{1}{2}\right)\Delta l, \left(k+\frac{1}{2}\right)\Delta l\right]$, *assuming that* $f_{\Delta t}(x)$ *is differentiable with respect to* x *on* $\left[\left(k-\frac{1}{2}\right)\Delta l, \left(k+\frac{1}{2}\right)\Delta l\right]$.

Next set $U^{MAX} = D^{MAX} = \infty$ *and* $\widetilde{\omega} = 0$. *Extend the domain of* $\widetilde{L}(t)$ *to* $[0,T]$ *by letting*

$$\widetilde{L}(t) = \widetilde{L}\left(\left\lfloor \frac{t}{\Delta t} \right\rfloor \Delta t\right), \quad t \in [0,T].$$

To constrain the asymptotic relationships amongst the parameters, for $n \in \mathbb{N}$ *let* $\Delta l = \Delta l(n)$, $\Delta t = \Delta t(n)$, $N = N(n)$, $u = u(n)$, $d = d(n)$, $\mathcal{K} = \mathcal{K}(n)$, $\widetilde{L}(t) =$

$\widetilde{L}(n,t)$, $\widetilde{S}(n,t) = S(0)\mathrm{e}^{\widetilde{L}(n,t)}$, *and*

$$\widetilde{\pi}(n,t) = \operatorname*{ess\,sup}_{\tau \in \mathcal{S}_{t,T}^{\widetilde{L}}} \mathrm{E}\left(\mathrm{e}^{-r(\tau-t)} h(\tau, \widetilde{S}(n,\tau)) \middle| \mathcal{F}_t^{\widetilde{L}}(n) \right), \quad t \in [0,T],$$

where $\mathbb{F}^{\widetilde{L}}(n) = \left(\mathcal{F}_t^{\widetilde{L}}(n)\right)_{t\in[0,T]}$ *is the natural filtration of* $\widetilde{L}(n,t)$ *and* $\mathcal{S}_{t,T}^{\widetilde{L}}$ *is the set of* $\mathbb{F}^{\widetilde{L}}(n)$*-stopping times in* $[t,T]$.

Take the setup given in Section 2.1 and the further assumption that the payoff function $h : [0,T] \times \mathbb{R}^+ \mapsto \mathbb{R}^+$ *is bounded and continuous. Suppose that, as* $n \to \infty$,

$$\frac{1}{\Delta t(n)} \int_{\infty}^{-d(n)\Delta l(n)} f_{\Delta t(n)}(x)\,\mathrm{d}x \to 0,$$
$$\frac{1}{\Delta t(n)} \int_{u(n)\Delta l(n)}^{\infty} f_{\Delta t(n)}(x)\,\mathrm{d}x \to 0,$$
$$\text{and} \quad \frac{\Delta l(n)}{\Delta t(n)} \to 0.$$

Theorem 3.5.1 then gives that

$$\lim_{n\to\infty} \widetilde{\pi}(n,0) = \pi(0).$$

Note that theoretical convergence does not guarantee computational tractability of the problem, but can guide parameter choice.

Asymptotic convergence of a lattice model for general, possibly infinite activity, Lévy processes was proved in Maller et al. (2006). The scheme used in Maller et al. (2006) is similar to that derived from the Lévy measure for jump-diffusion processes in Këllezi and Webber (2004). Under this scheme, Maller et al. (2006) used theoretical results from Mulinacci and Pratelli (1998) to establish convergence of the American price process in distribution for Meyer-Zheng topology. Convergence of processes in the Meyer-Zheng topology guarantees convergence of finite dimensional distributions on a set of Lebesgue measure one on $[0,T]$ (Meyer and Zheng 1984). The Meyer-Zheng topology is treated in more detail in Section 2.5. Using these

results, American options were successfully priced empirically under the Variance Gamma and Normal Inverse Gaussian models.

New theoretical results that became available in Coquet and Toldo (2007) allowed a stronger result for fundamentally the same model to be demonstrated in Szimayer and Maller (2007). For the sake of brevity, we describe the approximation scheme for a pure jump Lévy process, but it can be extended to accommodate a general Lévy process as described in Szimayer and Maller (2007). Assume the setup given in Section 2.1. The time interval $[0, T]$ is broken into $N(n)$, $n \in \mathbb{N}$, not necessarily equal partitions $\{t_0 = 0, t_1, \ldots, t_{N(n)} = T\}$. Take $\Delta t(n) = \max_{j \in \{1,\ldots,N(n)\}}(t_j - t_{j-1})$. Define a jump selection region $\mathcal{J}(n) = [-M(n), m(n)) \cup (m(n), M(n)]$, $m(n), M(n) \in \mathbb{R}^+$, and a jump size $\Delta(n) \in \mathbb{R}^+$. Assume that as $n \to \infty$, $\Delta t(n) \to 0$, $\Delta(n) \to 0$, $m(n) \to 0$, and $M(n) \to \infty$. To simplify the following definition, we assume that $M(n)/\Delta(n) \in \mathbb{N}$ and $m(n)/\Delta(n) \in \mathbb{N}$. This assumption is not made in Szimayer and Maller (2007) and isn't necessary for the theoretical development given there. For the approximating process $L(n, t)$, the transition probability mass function is then defined

$$
\begin{aligned}
&\mathrm{P}(\Delta L(n, t_j) = k\Delta(n)) \\
&= \begin{cases}
(1 - p_0(n)) \frac{\Pi\{(k\Delta(n), (k+1)\Delta(n))\}}{\Pi\{\mathcal{J}(n)\}}, & (k\Delta(n), (k+1)\Delta(n)) \subset \mathcal{J}(n) \cap (0, \infty), \\
p_0(n), & k = 0, \\
(1 - p_0(n)) \frac{\Pi\{((k-1)\Delta(n), k\Delta(n))\}}{\Pi\{\mathcal{J}(n)\}}, & ((k-1)\Delta(n), k\Delta(n)) \subset \mathcal{J}(n) \cap (-\infty, 0), \\
0, & \text{otherwise},
\end{cases}
\end{aligned}
$$

for $j \in \{1, \ldots, N(n)\}$ where $p_0(n) = \mathrm{e}^{-(t_j - t_{j-1})\Pi\{\mathcal{J}(n)\}}$. $L(n, t)$ is kept constant for $t \in [0, T] \setminus \{t_j\}_{j \in \{1,\ldots,N(n)\}}$. We also define the value functions

$$
v(n, t, x) = \sup_{\tau \in \mathcal{S}_{0, T-t}(n)} \mathrm{E}\left(h\left(\tau, x\mathrm{e}^{L(n,\tau)}\right)\right)
$$

and

$$v(t,x) = \sup_{\tau \in \mathcal{S}_{0,T-t}} \mathrm{E}\left((\tau, x\mathrm{e}^{L(\tau)})\right),$$

where $\mathcal{S}_{0,T-t}(n)$ and $\mathcal{S}_{0,T-t}$ are the sets of stopping times in $[0, T-t]$ with respect to the natural filtrations of $L(n,t)$ and $L(t)$ respectively. The main result of Szimayer and Maller (2007) is that if the payoff function h is bounded and continuous and if

$$\Delta t(n)\Pi^2\{(-\infty, -m(n)) \cup (m(n), \infty)\} \to 0 \quad \text{and}$$
$$\Delta(n)\Pi\{(-\infty, -m(n)) \cup (m(n), \infty)\} \to 0$$

as $n \to \infty$, then $\lim_{n\to\infty} v(n,t,x) = v(t,x)$ for all $(t,x) \in [0,T] \times \mathbb{R}^+$.

Key to the theoretical development in Maller et al. (2006), Szimayer and Maller (2007), and Chapter 3 is the fact that $L(n,t)$ can be constructed from the original process $L(t)$. This construction produces a useful relationship between the natural filtrations of $L(n,t)$ and $L(t)$ and can be used to prove convergence, in some sense, of $L(n,t)$ to $L(t)$. We will therefore briefly explain the mechanism involved. Using a variant of the Lévy-Ito decomposition (Sato 1999, Theorem 19.2, p. 120), we can write

$$L(t) = \gamma(n)t + L^{(1)}(n,t) + L^{(2)}(n,t) + L^{(3)}(n,t),$$

where

$$\gamma(n) = \gamma - \int_{m(n)<|x|\leq 1} x\Pi(\mathrm{d}x),$$
$$L^{(1)}(n,t) \overset{\text{a.s.}}{=} \lim_{\varepsilon \searrow 0} \left(\sum_{s\in(0,t]} \Delta L(s)\, \mathbf{1}(\varepsilon < |\Delta L(s)| \leq m(n)) - t \int_{\varepsilon<|x|\leq m(n)} x\Pi(\mathrm{d}x) \right),$$
$$L^{(2)}(n,t) = \sum_{s\in(0,t]} \Delta L(s)\, \mathbf{1}(|\Delta L(s)| > M(n)),$$

and

$$L^{(3)}(n,t) = \sum_{s\in(0,t]} \Delta L(s)\, \mathbf{1}_{\mathcal{J}(n)}(\Delta L(s)).$$

That is, $L^{(1)}(n,t)$ is a process that captures all the "small" jumps of $L(t)$, $L^{(2)}(n,t)$ captures the "large" jumps, and $L^{(3)}(n,t)$ is a process with jump sizes that only lie in the jump selection region. To illustrate the relationship between $L(t)$ and $L^{(3)}(n,t)$, Figure 2.1 shows an example Lévy process, and Figure 2.2 shows the same process with small and large jumps removed. We now introduce

$$\begin{aligned}
\tau_j(n) &= \inf\{t \in (t_{j-1}, t_j] : \Delta L^{(3)}(t) \neq 0\} \quad \text{and} \\
\theta(n) &= \{\tau_j(n)\}_{j\in\{1,\ldots,N(n)\}}
\end{aligned}$$

for $j \in \{1,\ldots,N(n)\}$, where we adopt the convention that $\inf\phi = \infty$, which is to say that $\tau_j(n)$ is the first time that $L^{(3)}(n,t)$ jumps in the $(j-1)$th interval. These structures will allow us to decompose $L^{(3)}(n,t)$ further as

$$\begin{aligned}
L^{(3)}(n,t) &= L^{(3,1)}(n,t) + L^{(3,2)}(n,t) \quad \text{where} \\
L^{(3,1)}(n,t) &= \sum_{s\in[0,t]\setminus\theta(n)} \Delta L^{(3)}(n,s) \quad \text{and} \\
L^{(3,2)}(n,t) &= \sum_{s\in[0,t]\cap\theta(n)} \Delta L^{(3)}(n,s).
\end{aligned}$$

$L^{(3,2)}(n,t)$ is illustrated in Figure 2.3, which shows the same process as that in Figure 2.2 with every jump but the first in each interval dropped. The final step is to adjust $L^{(3,2)}(n,t)$ so that it lies on a lattice in $[0,T]\times\mathbb{R}^+$, which is accomplished using

$$L(n,t) = \gamma(n)t + \left\lfloor L^{(3,2)}\left(n, \max\left([0,t]\cap\{t_j\}_{j\in\{1,\ldots,N(n)\}}\right)\right)/\Delta(n)\right\rfloor \Delta(n).$$

This step can be seen in Figure 2.4, which takes every jump in Figure 2.3 and shifts it onto a regular grid.

Another technique for calculating prices for American options that uses a dynamic programming approach and also bears a probabilistic interpretation is that of *Canadization*, a term coined in Carr (1998). Important to this technique are two theoretical financial instruments: the perpetual option, which is an American style option with infinite time to maturity, and the Canadian option, which is an American option where the time to maturity is a random variable. If the time to maturity follows an independent exponential distribution, then these two types of instruments share the property that the exercise boundary ceases to be a function of time and takes a constant value. This is a consequence of the memorylessness of the exponential distribution. In such cases the problem of valuing the option is simpler than for an American option with a finite time horizon. For instance, under Black-Scholes assumptions, the price of a Canadian put with exponentially distributed time to maturity can be found explicitly (Carr 1998). Expressions for the prices of perpetual options with certain payoffs and for a wide class of Lévy processes are given in Boyarchenko and Levendorskiĭ (2002) and Mordecki (2002) in terms of the Wiener-Hopf factorisation and overall supremum or infimum of the process in question, respectively.

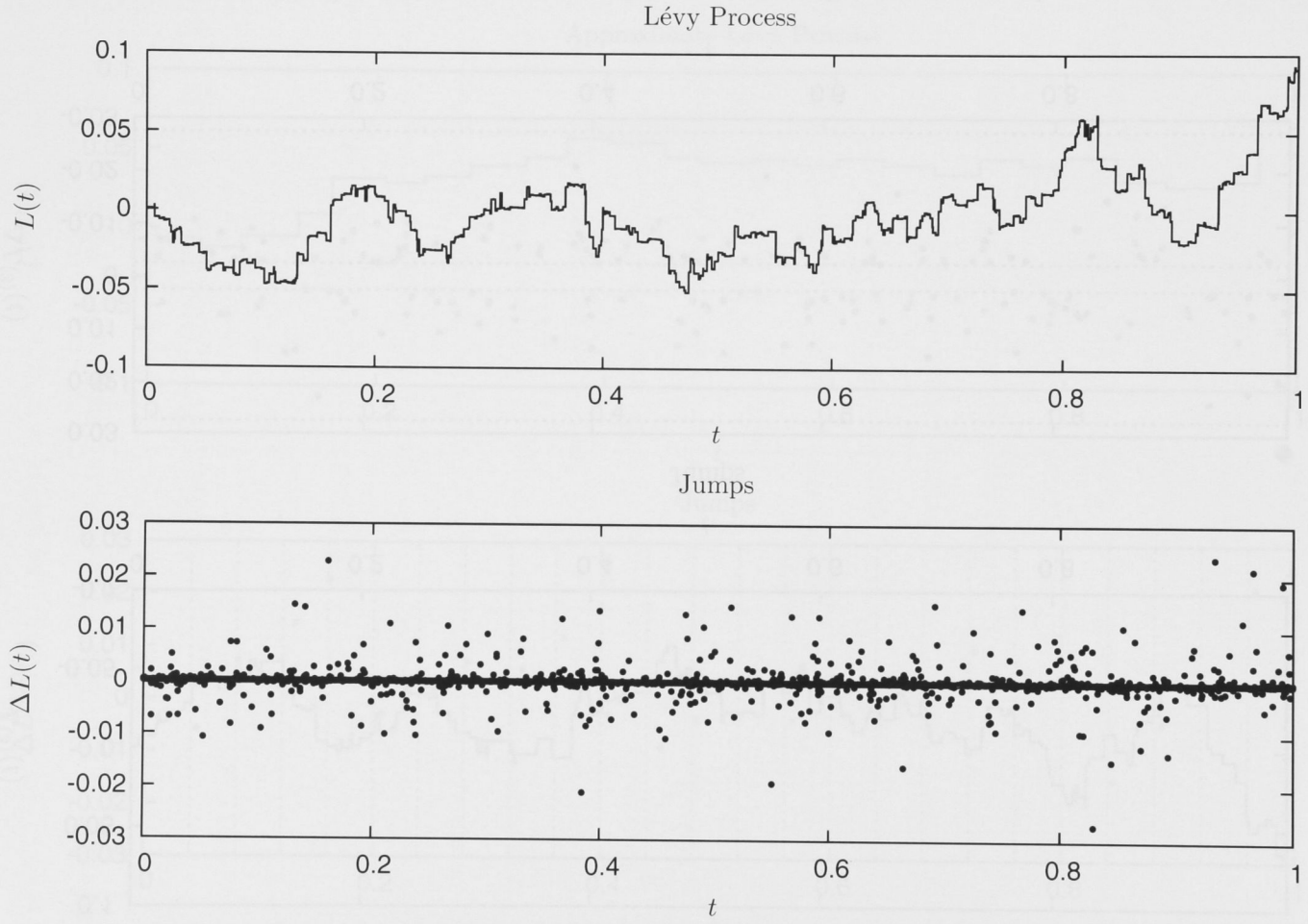

Figure 2.1: Original Lévy process. This example is a Variance Gamma ($b = 100$, $\mu = -0.14$, $\sigma = 0.12$) process, simulated using a series representation (Cont and Tankov 2004, Section 6.5).

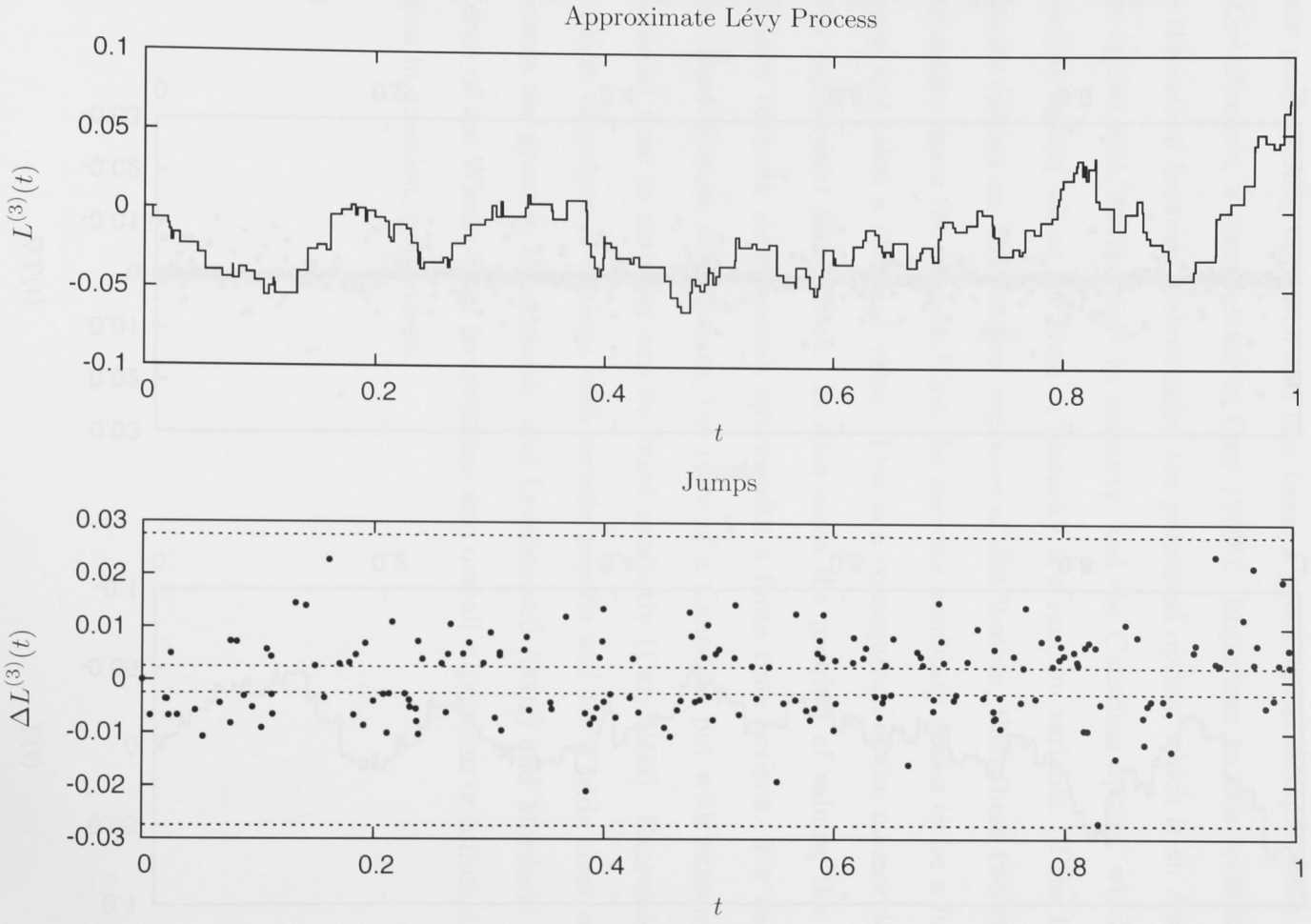

Figure 2.2: Lévy process with large and small jumps removed.

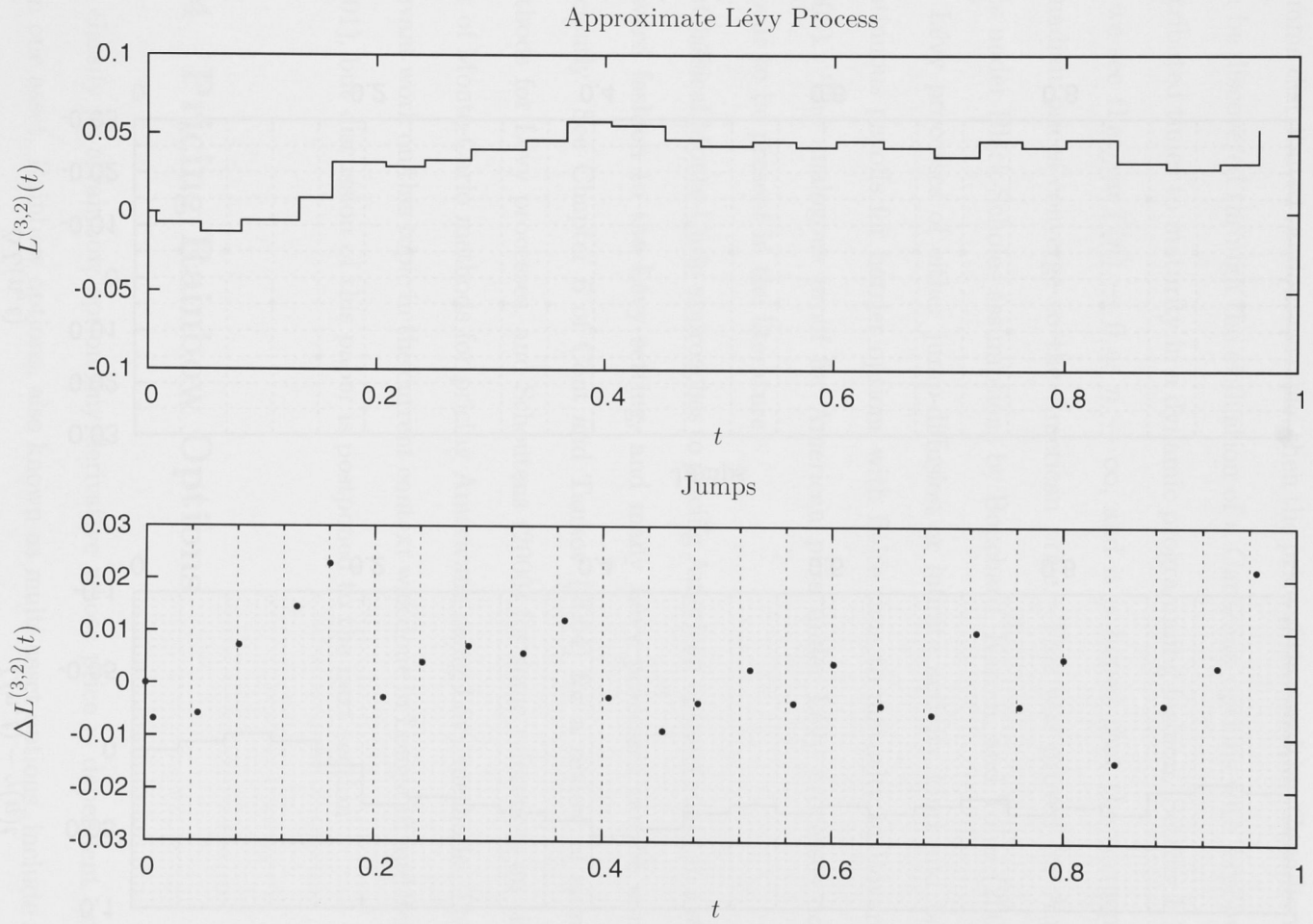

Figure 2.3: Lévy process with large and small jumps removed and all but one jump in each time interval dropped.

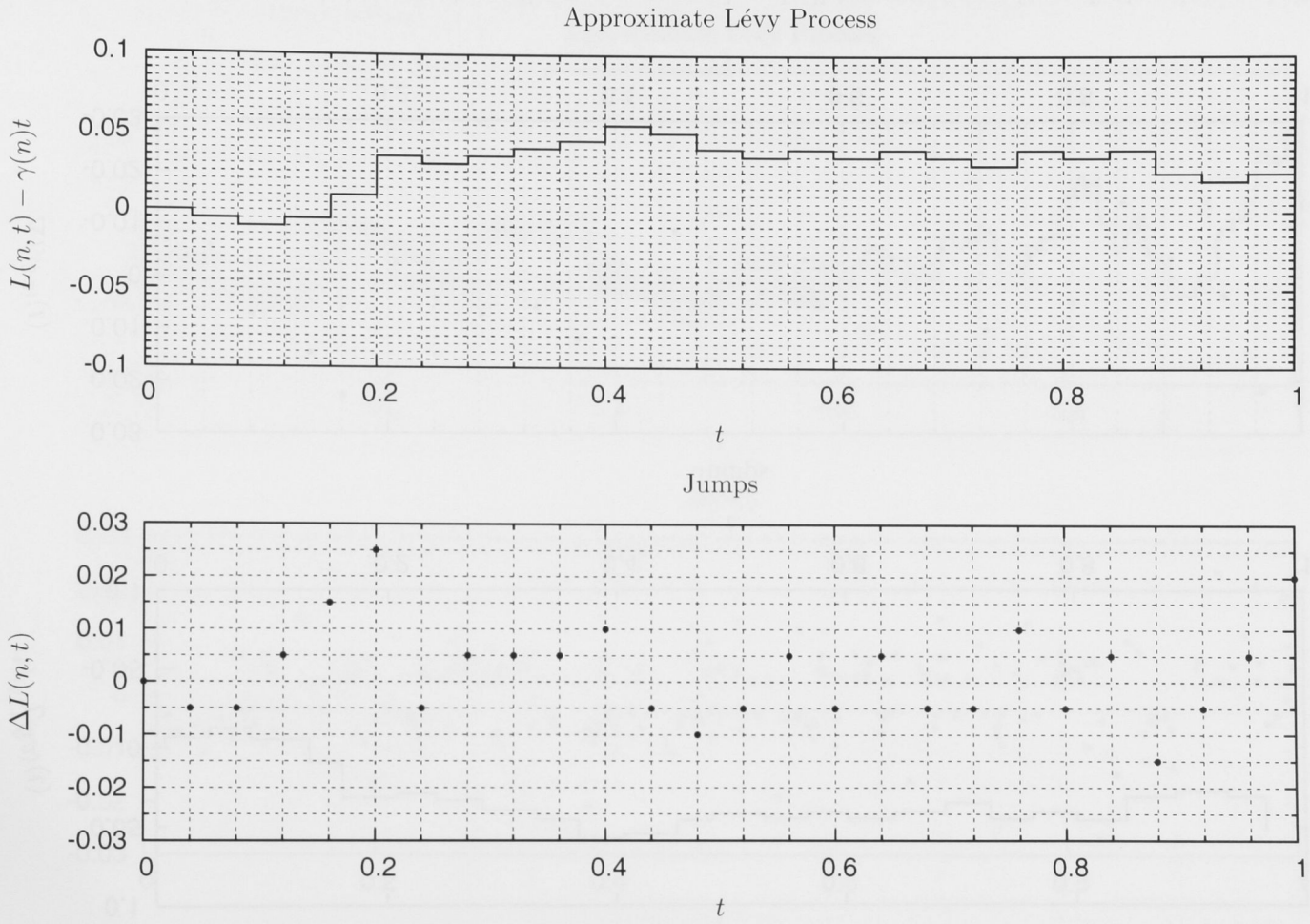

Figure 2.4: Lévy process with large and small jumps removed and all but one jump in each time interval dropped, with each jump moved onto a regular grid.

Key to Canadization is finding a sequence of Canadian prices that converge to the price of an American option. If we assume instead that the time to maturity $T(n) \in \mathbb{R}$ is the sum of $n \in \mathbb{N}$ iid exponential random variables, that is that $T(n)$ is a Gamma distributed random variable, then the price of the resulting Canadian option can be discovered through the evaluation of n Canadian options with exponentially distributed times to maturity in a dynamic programming fashion. Setting $\mathrm{E}\,T(n) = T$, we see that $\mathrm{var}\,T(n) \to 0$ as $n \to \infty$, and it is hoped that the corresponding Canadian prices converge to the American price. This was proved for American puts under Black-Scholes assumptions by Bouchard, Karoui, and Touzi (2005) and for Lévy processes of either jump-diffusion or infinite activity kind and bounded continuous payoffs for barrier options with finite time to maturity in Boyarchenko (2008). The analogous proof for American puts under Lévy processes does not appear to be present in the literature.

Classical Monte-Carlo approaches to pricing American options adapt in a straightforward fashion to the Lévy setting, and many Lévy processes can be simulated efficiently. See Chapter 6 of Cont and Tankov (2004) for a review of simulation methods for Lévy processes, and Schoutens (2006) for some references on the subject of Monte-Carlo methods for pricing Americans using Lévy methods. The most relevant work on this topic in the current context was done in Longstaff and Schwartz (2001), but discussion of this paper is postponed to the next section.

2.4 Pricing Rainbow Options

We classify as a rainbow option any derivative whose price is dependent on more than one asset. Rainbow options, also known as multi-asset options, include a wide variety of contingent claims. Some examples are basket options, spread options, options delivering the best of a number of assets and cash, quantos, which link foreign exchange and stock prices, agricultural futures contracts that allow delivery

of different grades of commodity, and bond futures that allow delivery of one of a predefined set of bonds. For more details on the different rainbow options available see Ong (1996), or see Briys, Bellalah, Mai, and de Varenne (1998, p. 313) for a description of various multivariate contracts.

For European two-colour rainbows, that is options on two underlying assets, a trick exists that allows explicit pricing in some cases. For instance, for spread options with payoff $h(T, S) = (S_1(T) - S_2(T))^+$, under Black-Scholes assumptions there is a closed form solution for the price known as the Margrabe formula. The trick is to change numeraire to one of the risky assets, that is to use the price of one asset as the accounting unit. For a full explanation in the Black-Scholes context see Hunziker and Koch-Medina (1996). The change of numeraire technique works equally well for (time-inhomogeneous) Lévy models (Eberlein and Papapantoleon 2005), although the price must be calculated using a numerical technique, such as a Fast Fourier Transform (FFT), to evaluate the implicit expectation. FFTs can also be used for pricing more general multivariate options, not including Americans, under some light constraints (Eberlein, Glau, and Papapantoleon 2008). These techniques for pricing European rainbow options using Lévy models could potentially be useful when pricing American rainbow options using numerical techniques by providing baselines for comparison.

In the Black-Scholes setting, the change of numeraire that yielded the Margrabe formula can be used to price American spread options using a one dimensional binomial lattice (Derman 1996). The binomial or Cox-Ross-Rubinstein (CRR) model is a particularly simple lattice model in which log prices follow a binomial process (Cox, Ross, and Rubinstein 1979). In general, however, multidimensional numerical methods are required for pricing rainbows. For two-colour rainbows, one solution is to use a binomial pyramid (Briys et al. 1998, p. 320–323). As its name suggests, the triangular lattice of the binomial model is replaced with a pyramid in the three dimensional space of the two risky asset prices and time. See Figure 2.5 for a depiction

of a three level pyramid lattice. The pyramid can be engineered to be recombining so that the n-th layer contains $(n+1)^2$ nodes, although the jump sizes and risk neutral probabilities are calculated in a different fashion to those for the standard CRR model.

A multinomial pyramid is used to evaluate a two-colour American Rainbow using a jump-diffusion model in Martzoukos (2003). The model consists of the sum of a correlated Gaussian process and several multivariate compound Poisson processes intended to model common market shocks. The multinomial pyramid is similar to the binomial pyramid, but jumps are assumed to reside on a square $M \times M$ grid for M an integer greater than two. The pyramid structure is a generalisation of the lattice model of Amin (1993), who modelled a univariate jump-diffusion process using a CRR model augmented by a larger regular grid on which transition probabilities were chosen to approximate the jump-size distribution. Martzoukos (2003) obtained generalisations of the Margrabe formula in this context and, using lattices calibrated from these analytic results, American prices.

A significantly more general, Monte-Carlo based algorithm is suggested in Longstaff and Schwartz (2001). It is intended to be used with "general stochastic processes", even as far as general semimartingales, in one or multiple dimensions, with the only restriction on payoff functions being that they have finite variance in the appropriate probabilistic setting. We state the algorithm, which is designated by its authors as least squares Monte-Carlo (LSM), as Algorithm 2.4.1, in which we assume the notation of Section 2.1, but not necessarily the model. For simplicity, we focus on the special case of a Markovian process and a vanilla American option, although LSM was originally framed for more general exotic options. Unlike the authors we also assume a constant interest rate.

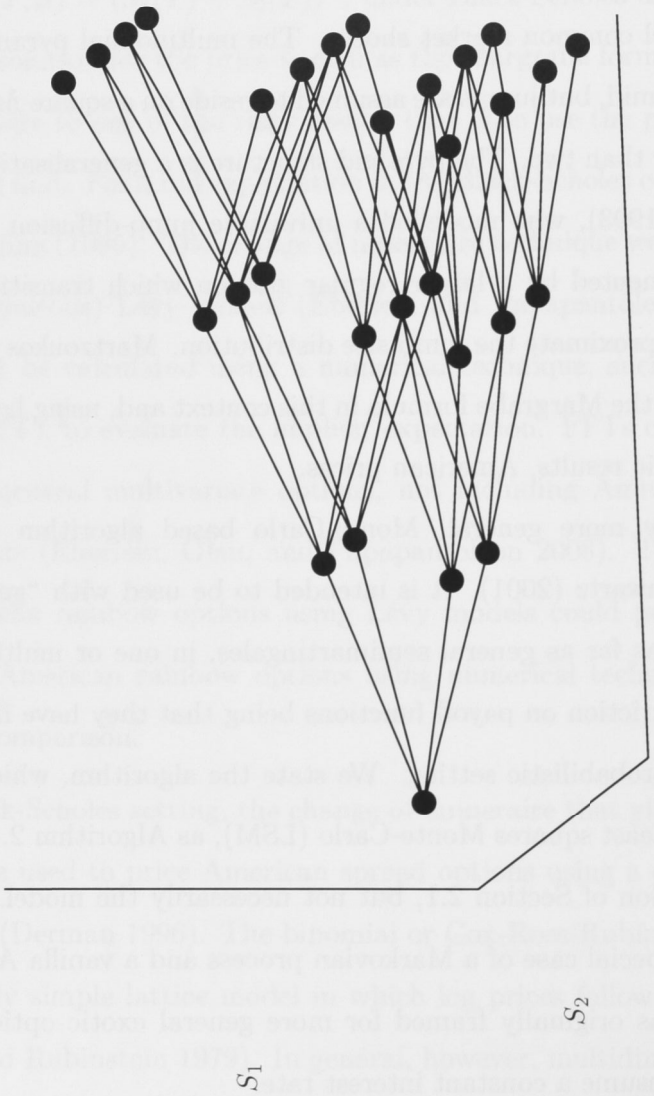

Figure 2.5: Diagrammatic representation of a binomial pyramid lattice as used for pricing two-colour American rainbows

LSM first attempts to discover the optimal exercise strategy for the option, then values it using this strategy. We denote its approximation to the optimal stopping strategy as $\widehat{\tau}$, which is a random variable defined for $N \in \mathbb{N}$ paths taken in a random sample. The definition of $\widehat{\tau}$ evolves as the algorithm works backwards through a time partition of $K \in \mathbb{N}$ steps. At each time step t_k, $k \in \{1, \ldots, K\}$, the information already gleaned about $\widehat{\tau}$ is used to build an approximation $\widehat{v}(t_k, \cdot) : \mathbb{R}^{+d} \mapsto \mathbb{R}$ to the conditional expectation of future cash flows as a function of underlying asset prices. This approximation takes the form of a truncated expansion in terms of basis functions $L_j : \mathbb{R}^{+d} \mapsto \mathbb{R}$, $j \in \mathbb{N}$, that span the set of square-integrable functions:

$$\widehat{v}(t_k, x) = \sum_{j=1}^{U} a_{j,k} L_j(x),$$

where $a_{j,k} \in \mathbb{R}$ and $U \in \mathbb{N}$. This expansion is found by the method of least squares. The possible bases mentioned in Longstaff and Schwartz (2001) include Laguerre, Hermite, Legendre, Chebyshev, Gegenbauer, and Jacobi polynomials, all of which are described in Abramowitz and Stegun (1964, Chapter 22), as well as Fourier and trigonometric sequences. Convergence of prices obtained using the LSM model to theoretical prices is proved for a simple example in Longstaff and Schwartz (2001), but a general proof is not provided.

Several examples are provided in Longstaff and Schwartz (2001), including an American put and an American-Bermuda-Asian option under Black-Scholes assumptions, an American put for a jump-diffusion process, and a best-of Bermudan option on five independent assets, also with Black-Scholes assumptions. Longstaff and Schwartz briefly discuss the issue of how the number of basis functions required in the expansion of $\widehat{v}(t_k, \cdot)$ might increase with the dimensionality of the underlying assets. Intuition might lead one to expect an exponential relationship, but it is observed empirically in Longstaff and Schwartz (2001) that the rate is much slower. This property has a theoretical basis in the field of nonlinear approximation, and is

Algorithm 2.4.1 The LSM algorithm for pricing an American option

Define the time partition $\{t_0 = 0, t_1, \ldots, t_K = T\}$

Draw a random sample of N paths $\mathcal{P} \subset \Omega$

$\widehat{\tau} \Leftarrow T$

for $k = K - 1$ to 0 **do**

$\mathcal{H}_k \Leftarrow \{\omega \in \mathcal{P} : h(t_k, S(t_k;\omega)) > 0\}$

Fit $\widehat{v}(t_k, \cdot)$ to $\left\{\left(S(t_k;\omega), \mathrm{e}^{-r(\widehat{\tau}(\omega)-t_k)} h(\widehat{\tau}(\omega), S(\widehat{\tau}(\omega);\omega))\right) : \omega \in \mathcal{H}_k\right\}$

for all $\omega \in \mathcal{H}_k$ **do**

if $\widehat{v}(t_k, S(t_k;\omega)) < h(t_k, S(t_k;\omega))$ **then**

$\widehat{\tau}(\omega) \Leftarrow t_k$

end if

end for

end for

return $\frac{1}{N} \sum_{\omega \in \mathcal{P}} \mathrm{e}^{-r\widehat{\tau}(\omega)} h(\widehat{\tau}(\omega), S(\widehat{\tau}(\omega);\omega))$

exploited by the next algorithm that we discuss.

In Longstaff and Schwartz (2001), the continuation values of an American option are approximated by series expansions for the purpose of determining the associated optimal stopping rule. In Kargin (2005), similar approximations are used to model the value function of the option directly for pricing. The motivation in this case is to price American options in a high-dimensional Black-Scholes setting using a lattice method. The naïve generalisation of the binomial pyramid discussed above quickly becomes computationally intractable in higher dimensions as the required number of nodes grows exponentially with dimension. However, methods drawn from the field of nonlinear approximation (DeVore 1998; Temlyakov 2003) can theoretically find approximations to certain classes of target functions such that the number of terms required to bring the properly specified approximation error under a specified threshold grows only polynomially with dimension. The Interpolative Lattice (IL) method described in Kargin (2005) therefore replaces each time-slice of nodes in the

lattice with such an approximation.

To understand the IL algorithm, assume a discretised version of the setup in Section 2.1 with the approximate process

$$\Delta\widetilde{L}(t) = \begin{cases} L(t_k) - L(t_{k-1}), & t = t_k, t_k \in \mathcal{T}, \\ 0, & t \in [0,T] \setminus \mathcal{T}, \end{cases}$$
$$\widetilde{L}(0) = 0,$$

where $\mathcal{T} = \{t_1 > 0, t_1, \ldots, t_K = T\}$, $K \in \mathbb{N}$, with corresponding underlying and option price processes $\widetilde{S}(t) = S(0)\mathrm{e}^{\widetilde{L}(t)}$ and

$$\widetilde{\pi}(t) = \operatorname*{ess\,sup}_{\tau \in \mathcal{S}^{\widetilde{L}}_{t,T}} \mathrm{E}\left(\mathrm{e}^{-r(\tau-t)} h(\tau, \widetilde{S}(\tau)) \middle| \mathcal{F}^{\widetilde{L}}_t \right), \quad t \in [0,T],$$

where $\mathbb{F}^{\widetilde{L}} = \left(\mathcal{F}^{\widetilde{L}}_t\right)_{t\in[0,T]}$ is the natural filtration of $\widetilde{L}(t)$ and $\mathcal{S}^{\widetilde{L}}_{t,T}$ is the set of $\mathbb{F}^{\widetilde{L}}$-stopping times in $[t,T]$. We define the value function

$$\widetilde{v}(t,x) = \sup_{\tau \in \mathcal{S}^{\widetilde{L}}_{t,T}} \mathrm{E}\left(\mathrm{e}^{-r(\tau-t)} h(\tau, \widetilde{S}(\tau)) \middle| \widetilde{S}(t) = x \right), \tag{2.4.1}$$

which can again be solved recursively for $t_k \in \mathcal{T}$ using the well-known Bellman equation (Wong 1996, p. 34):

$$\widetilde{v}(t_{k-1}, x) = \max\left(h(t_{k-1}, x), \mathrm{e}^{-r(t_k - t_{k-1})} \mathrm{E}\left[\widetilde{v}\left(t_k, \widetilde{S}(t_k)\right) \middle| \widetilde{S}(t_{k-1}) = x \right] \right), \tag{2.4.2}$$
$$\widetilde{v}(T,x) = h(T,x),$$

taking $t_0 = 0$. Note that $\widetilde{\pi}(t) = \widetilde{v}(t, \widetilde{S}(t))$, and that (2.3.3), which we encountered discussing the lattice models of Këllezi and Webber (2004), is a more specific form of (2.4.2).

The IL algorithm attempts to calculate $\widetilde{v}(t,0)$ numerically by building an ap-

proximation to $\widetilde{v}(t,\cdot)$ for each $t \in \mathcal{T}$, starting from $t = T$ and working recursively backward via the Bellman equation. For each $t \in \mathcal{T}$, this approximation takes the form

$$\widehat{v}(t,x) = \sum_{u=1}^{U(t)} \alpha_u(t)\phi\left(\langle b_u(t), \ln x\rangle + c_u(t)\right), \quad x \in \mathbb{R}^{+d}, \tag{2.4.3}$$

where for $u \in \{1,\ldots,U(t) \in \mathbb{N}\}$, $\alpha_u(t) \in \mathbb{R}^+$, $b_u(t) \in \mathbb{R}^d$, and $c_u(t) \in \mathbb{R}$, and $\phi : \mathbb{R} \mapsto \mathbb{R}$ is a so-called *ridge function*. The logarithm in (2.4.3) acts element-wise on its argument. The algorithm that fits $\widehat{v}(t_k,\cdot)$ to $\widetilde{v}(t_k,\cdot)$ for each $t_k \in \mathcal{T}$ works by evaluating

$$f(t_k,x) = \max\left(h(t_k,x), \mathrm{e}^{-r(t_{k+1}-t_k)}\,\mathrm{E}\left[\widehat{v}\left(t_{k+1}, S(t_{k+1})\right)|S(t_k) = x\right]\right),$$
$$f(T,x) = h(T,x),$$

for x on some finite grid $\mathcal{G} \subset \mathbb{R}^{+d}$, then fitting $\hat{v}(t_k,\cdot)$ to $\{(x, f(t_k,x))\}_{x\in\mathcal{G}}$ using progressively larger values of $U(t_k)$ until a goodness-of-fit criterion is satisfied. The final starting value $\widehat{v}(0,S(0)$ is evaluated as $f(t_0 = 0, S(0))$. This algorithm is described in very broad terms in Algorithm 2.4.2.

Algorithm 2.4.2 The IL algorithm for pricing Americans (Kargin 2005, Section 3).

1: Generate G, a grid – possibly irregular – in the factor space.

2: For each point $x \in G$, initialize the value function by computing the payoff at the final stage T.

3: Begin recursion over $t < T$: Compute an approximation to the value function at stage t using an approximation at stage $t+1$.

A ridge function is any suitable univariate function. Examples drawn from the nonlinear approximation literature include the sigmoid function (Barron 1993), sine function (Jones 1992), and hinging hyperplanes (Breiman 1993). The function ϕ chosen in Kargin (2005) is the multivariate Gaussian function. It is proved in Kargin

(2005) that a large class of multivariate functions can be approximated using this choice of ridge function to within an error δ using $\mathrm{O}(\delta^{-2})$ terms. A method for estimating upper and lower bounds for the true option value is also given.

Remark 2.4.1. *For the above setup, and also for the more general assumption that $L(t)$ is a Lévy process, Theorem 3.4.3 in Chapter 3 gives that*

$$\left(\widetilde{L}(t), h(t, \widetilde{S}(t)), \widetilde{\pi}(t)\right) \overset{\text{MZ}}{\Longrightarrow} (L(t), h(t, S(t)), \pi(t))$$

as $K \to \infty$, if h is bounded and continuous, $\mathcal{T} = \{k\Delta t\}_{k\in\{1,\ldots,K\}}$, and $\Delta t = T/K$. We use $\overset{\text{MZ}}{\Longrightarrow}$ to denote weak convergence in the Meyer-Zheng topology. This mode of convergence is explained in Section 2.5. The numerical convergence of $\widehat{v}(t,x)$ to $\widetilde{v}(t,x)$ is a separate issue which will be treated in Chapter 4.

Another lattice method is given in Lord, Fang, Bervoets, and Oosterlee (2007). This paper assumes a market model that includes all Lévy processes. The algorithm developed therein could be seen as a variant of Algorithm 2.4.2 above, although its motivation is very different. This motivation is the observation that for a one-dimensional underlying asset,

$$\mathscr{F}\left\{\mathrm{e}^{\alpha x}\,\mathrm{E}\left(\widetilde{v}\left(t_k, S(t_k)\right)|S(t_{k-1}) = \mathrm{e}^{x}\right)\right\}(u) = \mathscr{F}\left\{\mathrm{e}^{\alpha(x+z)}\widetilde{v}\left(t_k, \mathrm{e}^{x+z}\right)\right\}(u)\phi(-(u-\mathrm{i}\alpha))$$

for $\widetilde{v}$ defined in (2.4.1), where $\mathscr{F}$ denotes a Fourier transform

$$\mathscr{F}\{f(t)\}(u) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}ut} f(t)\,\mathrm{d}t,$$

$$\phi(x+y\mathrm{i}) = \mathrm{E}\left(\exp\left\{\mathrm{i}(x+y\mathrm{i})L(t_k - t_{k-1})\right\}\right) = \mathrm{e}^{-(t_k - t_{k-1})\Psi(x+y\mathrm{i})},$$

and $\mathrm{e}^{\alpha x}$, $\alpha > 0$, is a damping term required to make the Fourier transform converge. Thus, if the characteristic exponent Ψ is known, each iteration of the Bellman equation (2.4.2) can be calculated very quickly using an FFT and its inverse. However,

the FFT requires that the lattice be uniform in the log price dimension. Lord et al. demonstrate their algorithm by calculating several Bermudan and American prices for the Black-Scholes, VG, and CGMY models, and price a Bermudan on a basket put on four assets under Black-Scholes assumptions by generalising the algorithm to four dimensions. A proof of the convergence of the prices gotten via these lattice models to the continuous American price is not provided, but the errors introduced by discretisation in the price dimension are studied.

Remark 2.4.2. *Theorem 3.4.1 in Chapter 3 gives that for the setup in Remark 2.4.1 in one dimension, which is the lattice model used in Lord et al. (2007),*

$$\lim_{K\to\infty} \widetilde{\pi}(0) = \pi(0),$$

where once again h is bounded and continuous, $\mathcal{T} = \{k\Delta t\}_{k\in\{1,\dots,K\}}$, and $\Delta t = T/K$. The higher dimensional convergence result mentioned in Remark 2.4.1 also still holds. The convergence of their numerical implementation to their approximation model is shown in Lord et al. (2007).

It is possible to generalise the PIDE methods discussed in Section 2.3 to higher dimensional problems. According to Cont and Tankov (2004, Section 12.7), which endorses Monte-Carlo methods for pricing rainbows using Lévy processs, these techniques become inefficient for any problem involving more than two underlying assets. The problem of pricing two-colour American options using PIDEs is addressed in Clift and Forsyth (2008). This paper generalises the penalty method for solving an implicit version of the PIDE given in d'Halluin et al. (2004). The process assumed is the sum of a correlated Gaussian process and a bivariate compound Poisson process. Convergence of the method is proved under some criteria. Two Europeans and an American put on the minimum of two assets are priced, and a quadratic convergence rate is observed. It is also noted in Clift and Forsyth (2008) that the time slices of the exercise region $\mathcal{E}$ for the American option, which are two dimensional in this

case, are multiply connected.

As some algorithms, such as LSM, take advantage of the exercise region to cut down on computation time, we conclude this section with some remarks on the structure of this region for American rainbows. Hunziker and Koch-Medina (1996) warn that our intuition regarding the exercise region does not carry into higher dimensions, and that it may be optimal to exercise an American call on more than one asset, which is never true in the single asset case, assuming that the underlying asset does not pay dividends. Broadie and Detemple (2004) give several examples of this phenomenon in the Black-Scholes setting, including that for an American call on the best of two assets, that is $h(t, S_1, S_2) = (S_1 \vee S_2 - K)^+$, the exercise region, which is now a space in $[0, T] \times \mathbb{R}^{+2}$, takes the form

$$\begin{aligned}\mathcal{E} = &\{(t, S_1, S_2) \in [0, T] \times \mathbb{R}^{+2} : S_1 > S_2, S_1 \geq b_1(t, S_2)\} \\ &\cup \{(t, S_1, S_2) \in [0, T] \times \mathbb{R}^{+2} : S_1 < S_2, S_2 \geq b_2(t, S_1)\},\end{aligned}$$

where b_1 and b_2 are two boundary functions such that $S_2 < b_1(t, S_2)$ and $S_1 < b_2(t, S_1)$ for all $(t, S_1), (t, S_2) \in [0, T] \times \mathbb{R}^+$. Villeneuve (1999) expands on the examples given in Broadie and Detemple (2004) and provides several general theories, again under Black-Scholes assumptions.

2.5 Modes of Convergence

This section provides the definitions for several modes of convergence for stochastic processes and filtrations and some associated concepts that will be necessary for the theoretical development in Chapter 3. We start by defining the Skorokhod and Meyer-Zheng topologies.

Let $\mathbb{D}\left([0, T], \mathbb{R}^d\right)$ be the space of càdlàg $\mathbb{R}^d$-valued functions on $[0, T]$. The abbreviation càdlàg comes from the French for right-continuous with left-limits. The

Skorokhod J_1 distance between two processes X and Y in $\mathbb{D}\left([0,T],\mathbb{R}^d\right)$ is defined as

$$\rho(X,Y) = \inf_{\lambda\in\Lambda}\left\{\sup_{t\in[0,T]}\|X(t)-Y(\lambda(t))\| + \sup_{t\in[0,T]}|\lambda(t)-t|\right\},$$

where Λ is the set of strictly increasing continuous functions $\lambda : [0,T] \mapsto [0,T]$ with $\lambda(0) = 0$ and $\lambda(T) = T$. See Billingsley (1999, Chapter 3) for a full discussion of the Skorokhod J_1 topology. If a sequence of random processes $X(n,t) \in \mathbb{D}\left([0,T],\mathbb{R}^d\right)$, $n \in \mathbb{N}$, converges in distribution under the Skorokhod topology to a random process $X(t) \in \mathbb{D}\left([0,T],\mathbb{R}^d\right)$ (Prigent 2003, p. 69), then we write $X(n,t) \overset{\text{S}}{\Longrightarrow} X(t)$.

Let ℓ be the normalised Lebesgue measure on $[0,T]$ so that $\ell\{[0,T]\} = 1$. For any Borel function $g : [0,T] \mapsto \mathbb{R}$, the pseudo-path of g is the image measure of ℓ on $[0,T]\times[-\infty,\infty]$ under the mapping $t \mapsto (t, g(t))$. Denote by ψ the mapping from $\mathbb{D}([0,T],\mathbb{R})$ to the compact space of all probability measures on $[0,T]\times[-\infty,\infty]$ that maps a function g to its pseudo-path. ψ is injective and provides an imbedding of $\mathbb{D}([0,T],\mathbb{R})$ into the space of all probability measures on $[0,T]\times[-\infty,\infty]$. The induced topology on $\mathbb{D}([0,T],\mathbb{R})$ is then the Meyer-Zheng topology. For a summary of the properties of the Meyer-Zheng topology, see Mulinacci and Pratelli (1998), Meyer and Zheng (1984), or Prigent (2003, p. 74).

As an aid to intuition, if a series of functions $g_n(t) \in \mathbb{D}([0,T],\mathbb{R})$, $n \in \mathbb{N}$, converges to $g(t) \in \mathbb{D}([0,T],\mathbb{R})$ in the Meyer-Zheng topology, an equivalent condition is that for every bounded and continuous function $f : [0,T]\times[-\infty,\infty] \mapsto \mathbb{R}$,

$$\lim_{n\to\infty}\int_0^T f(s,g_n(s))\,\mathrm{d}s = \int_0^T f(s,g(s))\,\mathrm{d}s,$$

(Meyer and Zheng 1984, Lemma 1.).

The Meyer-Zheng topology on the product space $\mathbb{D}\left([0,T],\mathbb{R}^d\right)$ is the product topology. This implies that the family of random processes $X(n,t) \in \mathbb{D}\left([0,T],\mathbb{R}^d\right)$, $n \in \mathbb{N}$, is tight if each family of processes $\{X_j(n,t)\}_{n\in\mathbb{N}}$, $j \in \{1,\ldots,d\}$, is tight

(Mulinacci and Pratelli 1998). See, for example, (Prigent 2003, p. 70) for the definition of tightness. If a sequence of random processes $X(n,t) \in \mathbb{D}\left([0,T],\mathbb{R}^d\right)$, $n \in \mathbb{N}$, converges in distribution under the Meyer-Zheng topology to a random process $X(t) \in \mathbb{D}\left([0,T],\mathbb{R}^d\right)$, then we write $X(n,t) \overset{\text{MZ}}{\Longrightarrow} X(t)$.

We wish to employ results from Mulinacci and Pratelli (1998) and Coquet and Toldo (2007). Both papers use a condition known as Aldous' criterion for tightness, but each defines it slightly differently. In the current context we will say that a sequence of random processes $\{X(n,t)\}_{n\in\mathbb{N}}$ satisfies Aldous' criterion for *tightness in probability* if, for all $\varepsilon > 0$,

$$\lim_{\delta \searrow 0} \limsup_{n\to\infty} \sup_{\sigma,\tau\in\mathcal{S}^X_{0,T}(n);\sigma\leq\tau\leq\sigma+\delta} \mathrm{P}(\|X(n,\tau) - X(n,\sigma)\| \geq \varepsilon) = 0,$$

where $\mathcal{S}^X_{t,T}(n)$ is the set of $[t,T]$ valued $\mathbb{F}^{X(n)}$-stopping times and $\mathbb{F}^{X(n)}$ is the natural filtration of $X(n,t)$. Using the same notation, we will say that the sequence of processes satisfies Aldous' criterion for *tightness in expectation* if

$$\lim_{\delta \searrow 0} \limsup_{n\to\infty} \sup_{\sigma,\tau\in\mathcal{S}^X_{0,T}(n);\sigma\leq\tau\leq\sigma+\delta} \mathrm{E}\,\|X(n,\tau) - X(n,\sigma)\| = 0.$$

Consider a filtration $(\mathcal{F}_t)_{t\in[0,T]} = \mathbb{F}$ and a sequence of filtrations $(\mathcal{F}_t(n))_{t\in[0,T]} = \mathbb{F}(n)$ for $n \in \mathbb{N}$. We say that that $\mathbb{F}(n)$ converges weakly to $\mathbb{F}$, or $\mathbb{F}(n) \overset{\text{W}}{\to} \mathbb{F}$, if $\rho(\mathrm{E}(\mathbf{1}_A\,|\mathcal{F}_\cdot(n)), \mathrm{E}(\mathbf{1}_A\,|\mathcal{F}_\cdot)) \overset{\text{P}}{\to} 0$, where ρ is the Skorokhod J_1 metric, for all $A \in \mathcal{F}_T$. See Coquet, Mémin, and Słominski (2001) for an in depth treatment of weak convergence of filtrations. Also, we will say that $\mathbb{F}(n)$ is nested in $\mathbb{F}$, or $\mathbb{F}(n) \subset \mathbb{F}$, if $\mathcal{F}_t(n) \subset \mathcal{F}_t$ for all $t \in [0,T]$.

The following lemma is proved in Coquet et al. (2001) in Remark 1 part 1 and will be used in Chapter 3.

Lemma 2.5.1. *Given filtrations $(\mathcal{F}_t(n))_{t\in[0,T]} = \mathbb{F}(n)$, $n \in \mathbb{N}$, and $(\mathcal{F}_t)_{t\in[0,T]} = \mathbb{F}$. If $X \in \mathbb{R}$ is an $\mathcal{F}_T$-measurable, integrable random variable and $\mathbb{F}(n) \overset{\text{W}}{\to} \mathbb{F}$ as $n \to \infty$,*

then $\rho\left(\mathrm{E}(X|\mathcal{F}_\cdot(n)), \mathrm{E}(X|\mathcal{F}_\cdot)\right) \xrightarrow{\mathrm{P}} 0$.

The following condition is used extensively in Mulinacci and Pratelli (1998). If $Y(t)$ is a stochastic process, $\mathbb{F}^Y$ is the natural filtration of $Y(t)$, and $\mathbb{F}$ is a filtration to which $Y(t)$ is adapted, then we say that the pair $(Y(t), \mathbb{F}^Y)$ satisfies hypothesis (H) if every $\mathbb{F}^Y$-martingale is also an $\mathbb{F}$-martingale.

2.6 Convergence of American Option Prices

We import the following results from Mulinacci and Pratelli (1998) and Coquet and Toldo (2007).

Theorem 2.6.1 (Coquet and Toldo (2007), Theorem 5). *Take a process* $X(t)$ *and a sequence of processes* $\{X(n,t)\}_{n\in\mathbb{N}}$, *all drawn from* $\mathbb{D}([0,T],\mathbb{R})$. *Let the filtrations* $(\mathcal{F}_t^X)_{t\in[0,T]} = \mathbb{F}^X$ *and* $\left(\mathcal{F}_t^X(n)\right)_{t\in[0,T]} = \mathbb{F}^X(n)$ *for* $n \in \mathbb{N}$ *be the filtrations generated by* $X(t)$ *and* $\{X(n,t)\}_{n\in\mathbb{N}}$. *Define* $\mathcal{S}_{0,T}^X$ *and* $\mathcal{S}_{0,T}^X(n)$ *for* $n \in \mathbb{N}$ *to be the sets of* $\mathbb{F}^X$- *and* $\mathbb{F}^X(n)$*-stopping times. Assume that* $f : [0,T] \times \mathbb{R} \mapsto \mathbb{R}$ *is a continuous and bounded function. If*

(i) $\rho(X(n,\cdot), X(\cdot)) \xrightarrow{\mathrm{P}} 0$ *as* $n \to \infty$,

(ii) $\{X(n,t)\}_{n\in\mathbb{N}}$ *satisfies Aldous' criterion for tightness in probability, and*

(iii) $\mathbb{F}^X(n) \xrightarrow{\mathrm{W}} \mathbb{F}^X$ *as* $n \to \infty$ *or* $\mathbb{F}^X(n) \subset \mathbb{F}^X$ *for all* $n \in \mathbb{N}$,

then

$$\sup_{\tau\in\mathcal{S}_{0,T}^X(n)} \mathrm{E}(f(\tau, X(n,\tau))) \to \sup_{\tau\in\mathcal{S}_{0,T}^X} \mathrm{E}(f(\tau, X(\tau))), \quad n \to \infty.$$

Take a random process $H(t) \in \mathbb{D}([0,T],\mathbb{R}^+)$ such that $H(\tau)$ is uniformly integrable for every stopping time τ. The Snell envelope of $H(t)$, $J(t)$, is defined to be the smallest supermartingale such that for all $t \in [0,T]$, $J(t) \geq H(t)$. It is a well

known result (Mulinacci and Pratelli 1998) that for stopping times $\tau \in [0,T]$ and $\sigma \in [\tau,T]$,

$$J(\tau) = \operatorname*{ess\,sup}_{\sigma\in[\tau,T]} \mathrm{E}\left(H(\sigma)|\,\mathcal{F}_\tau\right),$$

that is if $H(t) = \mathrm{e}^{-rt}h(t,S(t))$, then $J(t) = \pi(t)$ from equation (2.1.2). Further, for any stopping time τ and $A \in \mathcal{F}_\tau$,

$$\int_A H(\tau)\,\mathrm{dP} = \sup_{\sigma\in[\tau,T]} \int_A H(\sigma)\,\mathrm{dP}\,. \tag{2.6.1}$$

We extend the notation $\overset{\text{MZ}}{\Longrightarrow}$ to include convergence in distribution for random variables defined on $\mathbb{D}\left([0,T],\mathbb{R}^d\right)\times[0,T]$ equipped with the product topology, where $\mathbb{D}\left([0,T],\mathbb{R}^d\right)$ is endowed with the Meyer-Zheng topology. For more details, and to see that the convergence statements below are well defined, see Mulinacci and Pratelli (1998, p. 316).

Theorem 2.6.2 (Mulinacci and Pratelli (1998), Theorem 3.5). *Let $X(t)$, $J(t)$, $Y(t)$, $X(n,t)$, and $J(n,t)$, $n \in \mathbb{N}$ be positive random processes in $\mathbb{D}([0,T],\mathbb{R})$ such that $J(t)$ and $J(n,t)$ are the Snell envelopes of $X(t)$ and $X(n,t)$ respectively. Define $\mathbb{F}^X = \left(\mathcal{F}_t^X\right)_{t\in[0,T]}$ and $\mathbb{F}^X(n) = \left(\mathcal{F}_t^X(n)\right)_{t\in[0,T]}$ as the natural filtrations of $X(t)$ and $X(n,t)$ and $\mathcal{S}_{0,T}^X(n)$ as the set of $\mathbb{F}^X(n)$-stopping times. If*

(i) $X(n,t) \overset{\text{S}}{\Longrightarrow} X(t)$,

(ii) for all $n \in \mathbb{N}$ and $\tau(n) \in \mathcal{S}_{0,T}^X(n)$, $X(n,\tau(n))$ is uniformly integrable,

(iii) $\{X(n,t)\}_{n\in\mathbb{N}}$ satisfies Aldous' criterion for tightness in expectation, and

(iv) for all $\tau(n) \in \mathcal{S}_{0,T}^X(n)$ and every process $(X(t),Y(t),\theta) \in \mathbb{D}([0,T],\mathbb{R}^2)\times[0,T]$ such that $(X(n,t),J(n,t),\tau(n)) \overset{\text{MZ}}{\Longrightarrow} (X(t),Y(t),\theta)$, $\left(X(t),\mathbb{F}^{X,Y,\theta}\right)$ satisfies (H), where $\mathbb{F}^{X,Y,\theta}$ is the smallest right continuous filtration to which (X,Y) is adapted and for which θ is an $\mathbb{F}^{X,Y,\theta}$-stopping time,

then $(X(n,t), J(n,t)) \overset{\text{MZ}}{\Longrightarrow} (X(t), J(t))$.

Verification of hypothesis (H) in the context of the last theorem can be difficult to ascertain. A sufficient condition for $\left(X(t), \mathbb{F}^{X,H,J,\theta}\right)$ to satisfy (H) is that $X(t)$ is Markovian with respect to $\mathbb{F}^{X,H,J,\theta}$ (Mulinacci and Pratelli 1998, p. 314). The following theorem can be useful in demonstrating this condition.

Theorem 2.6.3 (Mulinacci and Pratelli (1998), Theorem 4.1). *Define the random processes* $X(t), X(n,t) \in \mathbb{D}([0,T], \mathbb{R})$, $n \in \mathbb{N}$, *and* $J(n,t), J(t) \in \mathbb{D}([0,T], \mathbb{R})$ *and the stopping time* θ. *Say that* $\mathbb{F}^X(n) = \left(\mathcal{F}_t^X(n)\right)_{t\in[0,T]}$ *is the natural filtration of* $X(n,t)$, $\mathbb{F}^{X,J}(n) = \left(\mathcal{F}_t^{X,J}(n)\right)_{t\in[0,T]}$ *is the natural filtration of* $(X(n,t), J(n,t))$, $\tau(n)$ *is a* $\mathbb{F}^X(n)$*-stopping time, and that* $(X(n,t), J(n,t), \tau(n)) \overset{\text{MZ}}{\Longrightarrow} (X(t), J(t), \theta)$. *Also define* $\mathbb{F}^{X,J,\theta} = \left(\mathcal{F}_t^{X,J,\theta}\right)_{t\in[0,T]}$ *as the smallest right continuous filtration to which* $(X(t), J(t))$ *is adapted and for which* θ *is a* $\mathbb{F}^{X,J,\theta}$*-stopping time. If, for all* $n \in \mathbb{N}$ *and* $0 \leq s < t \leq T$, $(X(n,t) - X(n,s))$ *is independent of* $\mathcal{F}_s^X(n)$ *and* $\mathcal{F}_s^{X,J}(n) = \mathcal{F}_s^X(n)$, *then* $(X(t) - X(s))$ *is independent of* $\mathcal{F}_s^{X,J,\theta}$.

In Chapter 3, Theorems 2.6.2 and 2.6.3 will be generalised to a multidimensional setting so that the results in Maller et al. (2006) can be similarly generalised. A theoretical contribution relevant to the convergence of the lattice models that were presented in Sections 2.3 and 2.4 will also be made. Remarks 2.3.1, 2.4.1, and 2.4.2 have already alluded to this work. These developments will be made in the setting provided by Section 2.1 and will use the concepts given in Section 2.5.

Chapter 3

Approximation Models and Convergence of American Prices

3.1 Setup

We assume the market model described in Section 2.1. That is the market consists of a risk free asset and d risky assets modelled using an exponential Lévy process as in (2.1.1). The model is defined on a probability space $(\Omega, \mathbb{F}, \mathrm{P})$, where P is an equivalent martingale measure so that $\mathrm{e}^{-rt} S_j(t) = \mathrm{E}\left(\mathrm{e}^{-rT} S_j(T) | \mathcal{F}_t\right)$ for $j \in \{1, \ldots, d\}$ and $0 \leq t \leq T$. Equivalent conditions for this in terms of the Lévy triplet (γ, A, Π) are that

$$\int_{\mathbb{R}^d} \mathrm{e}^{x_j} \Pi(\mathrm{d}x) < \infty \quad \text{and} \quad r - \gamma_j - \frac{A_{jj}^2}{2} + \int_{\mathbb{R}^d} (1 - \mathrm{e}^{x_j} + x_j \mathbf{1}_D(x)) \, \Pi(\mathrm{d}x) = 0$$

for $j \in \{1 \ldots, d\}$ (Cont and Tankov 2004, Proposition 3.18). Under these assumptions, the price of an American option with payoff $h : [0, T] \times \mathbb{R}^{+d} \mapsto \mathbb{R}^+$ follows the process defined in (2.1.2).

We introduce the notation $\mathbb{F}^L = \left(\mathcal{F}_t^L\right)_{t \in [0,T]}$ to indicate the natural filtration of

$L(t)$ and decompose $L(t)$ as

$$L(t) = \gamma t + B(t) + L^{\text{PJ}}(t), \tag{3.1.1}$$

where $B(t)$ is a Gaussian process on $\mathbb{R}^d$ with covariance matrix A and $L^{\text{PJ}}(t)$ is the pure jump process on $\mathbb{R}^d$ with Lévy measure Π.

3.2 The First Jump Approximation Model

First partition $[0,T]$ into $N(n)$ intervals of length $\Delta t(n) = \frac{T}{N(n)}$, where $N : \mathbb{N} \mapsto \mathbb{N}$ and is strictly increasing, and set $G(n) = \{0, \Delta t(n), \ldots, N(n)\Delta t(n) = T\}$ to be the time grid on which our calculations will take place.

Define a sequence of processes $\{L(n,t)\}_{n\in\mathbb{N}}$, where $t \in [0,T]$ and $T < \infty$,

$$L(n,t) = \gamma(n)t + B(n,t) + L^{\text{PJ}}(n,t), \tag{3.2.1}$$

with $\gamma(n) \in \mathbb{R}^d$ defined in such a manner that

$$\mathrm{E}\left(\mathrm{e}^{L_j(n,\Delta t(n))}\right) = \mathrm{e}^{r\Delta t(n)}, \quad j \in \{1,\ldots,d\}. \tag{3.2.2}$$

$B(n,t)$ is an approximation to $B(t)$ defined by

$$B(n,t) = B\left(\left\lfloor \frac{t}{\Delta t(n)} \right\rfloor \Delta t(n)\right), \quad t \in [0,T],$$

and $L^{\text{PJ}}(n,t)$ is an approximation to $L^{\text{PJ}}(t)$. It is defined as

$$L^{\text{PJ}}(n,t) = \sum_{i=1}^{\left\lfloor \frac{t}{\Delta t(n)} \right\rfloor} X(n,i), \tag{3.2.3}$$

where $\{X(n,i)\}_{i\in\{1,\ldots,N(n)\}}$ is a series of iid $\mathbb{R}^d$ valued random variables with proba-

bility measure

$$\mathrm{P}^{X(n)}(\mathcal{B}) = \frac{1 - \mathrm{e}^{-\Delta t(n)\Pi\{\mathcal{J}(n)\}}}{\Pi\{\mathcal{J}(n)\}}\Pi\{\mathcal{B} \cap \mathcal{J}(n)\} + \mathrm{e}^{-\Delta t(n)\Pi\{\mathcal{J}(n)\}}\delta_0(\mathcal{B}) \tag{3.2.4}$$

for Borel $\mathcal{B} \subset \mathbb{R}^d$, and δ_0 is Dirac measure at $0 \in \mathbb{R}^d$. Note that, while only the distribution of each $X(n,i)$ is given here, a pathwise relationship between $L^{\mathrm{PJ}}(n,t)$ and $L^{\mathrm{PJ}}(t)$ consistent with this definition will be given below.

We call $\mathcal{J}(n)$ the jump selection region and define it as follows. Let $\{\mathcal{R}(q)\}_{q\in\mathbb{R}^+}$ be a nested family of closed Borel sets on $\mathbb{R}^d$ satisfying $0 \in \mathcal{R}(q)$ and $\mathcal{R}(q) \subset \mathcal{R}(q')$ for $0 < q < q'$, and such that $\mathcal{R}(q) \searrow \{0\}$ as $q \searrow 0$ and $\mathcal{R}(q) \nearrow \mathbb{R}^d$ as $q \nearrow \infty$. Define two real sequences $\{m(n)\}_{n\in\mathbb{N}}$ and $\{M(n)\}_{n\in\mathbb{N}}$ such that $m(n) \searrow 0$ and $M(n) \nearrow \infty$ as $n \to \infty$. Then set $\mathcal{J}(n) = \mathcal{R}(M(n)) \setminus \mathcal{R}(m(n))$.

Correspondingly, we have the approximate underlying price process,

$$S(n,t) = \left(S_j(0)\mathrm{e}^{L_1(n,t)}, \ldots, S_d(0)\mathrm{e}^{L_d(n,t)}\right), \quad t \in [0,T] \tag{3.2.5}$$

and, outside the exercise region, the option price process,

$$\pi(n,t) = \operatorname*{ess\,sup}_{\tau\in\mathcal{S}_{t,T}(n)} \mathrm{E}\left(\mathrm{e}^{-r(\tau-t)}h(\tau, S(n,\tau))\middle|\, \mathcal{F}_t^L(n)\right). \tag{3.2.6}$$

Here $\mathcal{S}_{t,T}(n)$ is the set of $\mathbb{F}^L(n)$-stopping times in $[t,T]$ and $\mathbb{F}^L(n) = \{\mathcal{F}_t^L(n)\}_{t\geq 0}$ is the filtration induced on Ω by $L(n,t)$.

We call $\{L(n,t)\}_{n\in\mathbb{N}}$, $t \in [0,T]$, the sequence of first jump approximations to $L(t)$. $L(n,t)$ converges to $L(t)$ in the following fashion:

Theorem 3.2.1. *Assume that*

$$\lim_{n\to\infty} \Delta t(n)\Pi^2\left(\mathbb{R}^d \setminus \mathcal{R}(m(n))\right) = 0. \tag{3.2.7}$$

Then as $n \to \infty$,

(i) $\rho(L(n,\cdot), L(\cdot)) \xrightarrow{\text{P}} 0$,

(ii) $\mathbb{F}^L(n) \xrightarrow{\text{W}} \mathbb{F}^L$, *and*

(iii) $\{L(n,t)\}_{n\in\mathbb{N}}$ *fulfills Aldous' criterion for tightness in probability.*

Theorem 3.2.1(i) and (iii) are generalisations of Theorem 3.2(i) and (ii) in Szimayer and Maller (2007), in that Theorem 3.2(i) and (ii) in Szimayer and Maller (2007) prove the same result for a one dimensional first jump approximation model. Note, however, that the model assumed here is slightly different. In Szimayer and Maller (2007), $\Delta L(n,t)$ is binned to create a discrete random process, which in turn requires a more complicated definition of $B(n,t)$ than that shown above. Theorem 3.2.1(ii) is a new result.

The proof of Theorem 3.2.1 requires the following construction. Consider the processes

$$L^{\text{J}}(t) = \gamma t + L^{\text{PJ}}(t) \quad \text{and} \tag{3.2.8}$$

$$L^{\text{J}}(n,t) = \gamma(n)t + L^{\text{PJ}}(n,t), \tag{3.2.9}$$

where $L^{\text{PJ}}(t)$ is defined in (3.1.1) and $\gamma(n)$ and $L^{\text{PJ}}(n,t)$ are defined in (3.2.1). Using a variant of the Lévy-Ito decomposition (Sato 1999, Theorem 19.2, p. 120), for $n \in \mathbb{N}$ and $t \in [0,T]$, decompose $L^{\text{J}}(t)$ as

$$L^{\text{J}}(t) = \gamma^{\text{J}}(n)t + L^{(1)}(n,t) + L^{(2)}(n,t) + L^{(3)}(n,t), \tag{3.2.10}$$

where

$$\gamma^{\text{J}}(n) = \gamma - \int_{D\setminus\mathcal{R}(m(n))} x\Pi(\mathrm{d}x),$$

$$L^{(1)}(n,t) \overset{\text{a.s.}}{=} \lim_{\varepsilon\searrow 0}\left(\sum_{s\in(0,t]} \Delta L(s)\,\mathbf{1}_{\mathcal{R}(m(n))\setminus\mathcal{R}(\varepsilon)}(\Delta L(s)) - t\int_{\mathcal{R}(m(n))\setminus\mathcal{R}(\varepsilon)} x\Pi(\mathrm{d}x)\right),$$

$$L^{(2)}(n,t) = \sum_{s\in(0,t]} \Delta L(s)\,\mathbf{1}_{\mathbb{R}^d\setminus\mathcal{R}(M(n))}(\Delta L(s)), \quad \text{and}$$
$$L^{(3)}(n,t) = \sum_{s\in(0,t]} \Delta L(s)\,\mathbf{1}_{\mathcal{J}(n)}(\Delta L(s)).$$

Recall that $D = \{x : \|x\| \le 1\}$. We capture the first jump in our jump selection region for each $k \in \{1,\dots,N(n)\}$ with the construction

$$\Theta_k(n) = \{s \in ((k-1)\Delta t(n), k\Delta t(n)] : \Delta L(s) \in \mathcal{J}(n)\} \quad \text{and}$$
$$\tau_k(n) = \inf \Theta_k(n), \tag{3.2.11}$$

where we adopt the convention that $\inf\phi = \infty$. $L^{(3)}(n,t)$ can now be decomposed as

$$L^{(3)}(n,t) = L^{(3,1)}(n,t) + L^{(3,2)}(n,t), \tag{3.2.12}$$

where

$$L^{(3,2)}(n,t) = \sum_{k=1}^{\tau_k(n)\le t} \Delta L^{(3)}(\tau_k(n)) \quad \text{and}$$
$$L^{(3,1)}(n,t) = L^{(3)}(n,t) - L^{(3,2)}(n,t).$$

Note that, consistent with the construction of $L(n,t)$ via $L^{\text{PJ}}(n,t)$, which is defined in (3.2.4), we can specify

$$X(n,k) = \begin{cases} \Delta L^{(3)}(\tau_k(n)), & \tau_k(n) < \infty, \\ 0, & \text{otherwise,} \end{cases}$$

so that

$$L^{\text{PJ}}(n,t) = L^{(3,2)}\left(n, \left\lfloor \frac{t}{\Delta t(n)} \right\rfloor \Delta t(n)\right). \tag{3.2.13}$$

We proceed to prove Theorem 3.2.1 via a series of lemmas.

It will be shown that $L^{(1)}$, $L^{(2)}$, and $L^{(3,1)}$ are negligible for large n, so with that in mind we define a process with at most a single jump between nodes of $G(n)$:

$$L^{\mathrm{S}}(n,t) = \gamma^{\mathrm{J}}(n)t + L^{(3,2)}(n,t). \tag{3.2.14}$$

Lemma 3.2.1. *For $L^{(1)}(n,t)$ and $L^{(2)}(n,t)$ defined in* (3.2.10)*, as $n \to \infty$,*

$$\sup_{t\in[0,T]} \left\| L^{(1)}(n,t) \right\| \xrightarrow{\mathrm{P}} 0 \quad \textit{and} \quad \sup_{t\in[0,T]} \left\| L^{(2)}(n,t) \right\| \xrightarrow{\mathrm{P}} 0.$$

Lemma 3.2.1 generalises Lemma 6.1(a)(ii) and (b)(i) of Szimayer and Maller (2007), which present almost sure limits as $n \to \infty$ for the same quantities in one dimension. Proof of Lemma 3.2.1 initially follows similar lines to that of Lemma 6.1(a)(ii) in Szimayer and Maller (2007), but takes advantage of the fact that we only need convergence in probability here to present a much simpler proof than the corresponding proof for Lemma 6.1(b)(i).

Proof. Note that $L^{(1)}(n,t)$ is a martingale and a Lévy process with Lévy measure $\Pi^{(1)}(B) = \Pi(B \cap \mathcal{R}(m(n)))$ for Borel set B. Take $j \in \{1,\ldots,d\}$. Doob's maximal quadratic inequality gives that

$$\mathrm{E}\left(\sup_{t\in[0,T]} \left| L_j^{(1)}(n,t) \right| \right)^2 \leq 4\,\mathrm{E}\left(\left| L_j^{(1)}(n,T) \right|^2 \right).$$

Using Example 25.12 in Sato (1999, p. 163), we see that

$$\mathrm{E}\left| L_j^{(1)}(n,T) \right|^2 = T \int_{\mathcal{R}(m(n))} x_j^2 \Pi(\mathrm{d}x) \to 0, \quad n \to \infty,$$

which implies that $\sup_{t\in[0,T]}\left|L_j^{(1)}(n,t)\right| \xrightarrow{\mathrm{P}} 0$. The triangle inequality then gives

$$\sup_{t\in[0,T]}\left\|L^{(1)}(n,t)\right\| \leq \sum_{j=1}^{d}\sup_{t\in[0,T]}\left|L_j^{(1)}(n,t)\right| \xrightarrow{\mathrm{P}} 0, \quad n\to\infty.$$

For all $\varepsilon > 0$,

$$\begin{aligned}\mathrm{P}\left(\sup_{t\in[0,T]}\left\|L^{(2)}(n,t)\right\| > \varepsilon\right) &\leq \mathrm{P}\left(\sum_{t\in[0,T]}\mathbf{1}_{\mathbb{R}^d\setminus\mathcal{R}(M(n))}(\Delta L(t)) \neq 0\right)\\ &= 1-\exp\left\{-T\Pi\left(\mathbb{R}^d\setminus\mathcal{R}(M(n))\right)\right\},\end{aligned}$$

which tends to zero as $n\to\infty$ because, by definition, $\mathcal{R}(M(n))\to\mathbb{R}^d$ as $n\to\infty$. Therefore $\sup_{t\in[0,T]}\left\|L^{(2)}(n,t)\right\| \xrightarrow{\mathrm{P}} 0$ as $n\to\infty$. □

Lemma 3.2.2. *Under assumption* (3.2.7), *for* $L^{(3,1)}(n,t)$ *defined in* (3.2.12), *as* $n\to\infty$,

$$\sup_{t\in[0,T]}\left\|L^{(3,1)}(n,t)\right\| \xrightarrow{\mathrm{P}} 0.$$

Lemma 3.2.2 generalises Lemma 6.2 in Szimayer and Maller (2007) to arbitrary dimensions, and its proof proceeds similarly.

Proof. First define

$$\Theta_k'(n) = \Theta_k(n)\setminus\{\tau_k(n)\}$$

for $\tau_k(n)$ defined in (3.2.11) and $k\in\{1,\ldots,N(n)\}$. Using the triangle inequality, we can write

$$\sup_{i\in[0,T]}\left\|L^{(3,1)}(n,t)\right\| \leq \sum_{k=1}^{N(n)}\sum_{s\in\Theta_k'(n)}\|\Delta L(s)\|.$$

Let $\#\Theta_k(n)$ be the number of elements in $\Theta_k(n)$. The quantities $\{\#\Theta_k(n)\}_{k\in\{1,\ldots,N(n)\}}$

then form a series of iid Poisson distributed random variables with intensity parameter $\ell(n) = \Delta t(n)\Pi(\mathcal{J}(n))$. Now,

$$\begin{aligned}
&\mathrm{P}\left(\max_{k\in\{1,\ldots,N(n)\}} \sum_{s\in\Theta'_k(n)} \|\Delta L(s)\| > 0\right) \\
&\quad\le \sum_{k=1}^{N(n)} \mathrm{P}(\#\Theta_k(n) > 1) \le N(n)\mathrm{e}^{-\ell(n)} \sum_{k=2}^{\infty} \frac{\ell^k(n)}{k!} \\
&\quad\le N(n)(\Delta t(n)\Pi(\mathcal{J}(n)))^2 = T\Delta(t)\Pi^2(\mathcal{J}(n)) \\
&\quad\le T\Delta(t)\Pi^2(\mathbb{R}^d \setminus \mathcal{R}(m(n))).
\end{aligned}$$

This expression tends to 0 as $n \to \infty$ under assumption (3.2.7), so we have convergence in probability. □

Lemma 3.2.3. *Under assumption* (3.2.7), *with* $\gamma(n)$ *and* $\gamma^{\jmath}(n)$ *as defined in* (3.2.1) *and* (3.2.10), $\|\gamma^{\jmath}(n) - \gamma(n)\| \to 0$ *as* $n \to \infty$.

Remark 3.2.1. *When constructing* $L(n,t)$, *we could have assumed that* $\|\gamma^{\jmath}(n) - \gamma(n)\| \to 0$ *as* $n \to \infty$, *which is more general than choosing a particular* $\gamma(n)$ *with that property, as we have in* (3.2.2). *However, we prefer the more constructive nature of an explicit choice.*

Proof. From the definitions of $\gamma(n)$ and $\gamma^{\jmath}(n)$ we derive that for $j \in \{1,\ldots,d\}$,

$$\begin{aligned}
\gamma_j^{\jmath}(n) &= r - \frac{\sigma_j^2}{2} - \int_{\mathbb{R}^d} \left(\mathrm{e}^{x_j} - 1 - x_j\, \mathbf{1}_{\mathcal{R}(m(n))}(x)\right) \Pi(\mathrm{d}x), \quad \text{and} \\
\gamma_j(n) &= r - \frac{\sigma_j^2}{2} \\
&\quad - \frac{1}{\Delta t(n)} \ln\left\{\mathrm{e}^{-\Delta t(n)\Pi(\mathcal{J}(n))} + \frac{1 - \mathrm{e}^{-\Delta t(n)\Pi(\mathcal{J}(n))}}{\Pi(\mathcal{J}(n))} \int_{\mathcal{J}(n)} \mathrm{e}^{x_j} \Pi(\mathrm{d}x)\right\},
\end{aligned}$$

where σ_j^2 is the jth diagonal element of the Gaussian covariance matrix A. To

compare $\gamma_j^{\mathrm{J}}(n)$ with $\gamma_j(n)$ we expand the former using the Taylor series

$$\ln x = -\sum_{i=1}^{\infty} \frac{1}{i}(1-x)^i, \quad x \in (0,2].$$

The following working shows that the logarithm function is analytic for the arguments we consider. There exists a c such that $\mathcal{R}(c) \subset \{x : \|x\| \leq 1\}$ and, as $\mathrm{E}\, S_j(t) = S_j(0)\mathrm{e}^{rt}$, our Lévy process possesses exponential moments, so for all n

$$\int_{\mathcal{J}(n)} \mathrm{e}^{x_j} \frac{\Pi(\mathrm{d}x)}{\Pi(\mathcal{J}(n))} \leq \mathrm{e}\frac{\Pi(\mathcal{R}(c) \cap \mathcal{J}(n))}{\Pi(\mathcal{J}(n))} + \int_{\mathcal{J}(n)\setminus\mathcal{R}(c)} \mathrm{e}^{x_j} \frac{\Pi(\mathrm{d}x)}{\Pi(\mathcal{J}(n))} < \infty.$$

By Assumption (3.2.7), we also have that

$$\Delta t(n)\Pi(\mathcal{J}(n)) \to 0 \quad \text{as} \quad n \to \infty. \tag{3.2.15}$$

So, for sufficiently large n, Taylor expansions of the logarithm and exponential give that

$$\begin{aligned} &\ln\left\{\mathrm{e}^{-\Delta t(n)\Pi(\mathcal{J}(n))} + \frac{1-\mathrm{e}^{-\Delta t(n)\Pi(\mathcal{J}(n))}}{\Pi(\mathcal{J}(n))}\int_{\mathcal{J}(n)} \mathrm{e}^{x_j}\Pi(\mathrm{d}x)\right\} \\ &\quad = \Delta t(n)\Pi(\mathcal{J}(n))\left\{-1 + \int_{\mathcal{J}(n)} \mathrm{e}^{x_j}\frac{\Pi(dx)}{\Pi(\mathcal{J}(n))}\right\} + \mathrm{O}\left((\Delta t(n)\Pi(\mathcal{J}(n)))^2\right). \end{aligned} \tag{3.2.16}$$

Thus,

$$\begin{aligned} \gamma_j^{\mathrm{J}}(n) - \gamma_j(n) &= \int_{\mathcal{J}(n)} (\mathrm{e}^{x_j} - 1)\,\Pi(\mathrm{d}x) - \int_{\mathbb{R}^d} \left(\mathrm{e}^{x_j} - 1 - x_j\,\mathbf{1}_{\mathcal{R}(m(n))}(x)\right)\Pi(\mathrm{d}x) \\ &\quad + \mathrm{O}\left(\Delta t(n)\Pi^2(\mathcal{J}(n))\right) \\ &= -\int_{\mathbb{R}^d\setminus\mathcal{R}(M(n))} (\mathrm{e}^{x_j} - 1)\,\Pi(\mathrm{d}x) - \int_{\mathcal{R}(m(n))} (\mathrm{e}^{x_j} - 1 - x_j)\,\Pi(\mathrm{d}x) \\ &\quad + \mathrm{O}\left(\Delta t(n)\Pi^2(\mathcal{J}(n))\right). \end{aligned}$$

Now, $\mathcal{R}(M(n)) \to \mathbb{R}^d$, $\mathcal{R}(m(n)) \to 0$, and $\Delta t(n)\Pi^2(\mathcal{J}(n)) \to 0$ by assumption, so

$$\|\gamma^{\mathrm{J}}(n) - \gamma(n)\| \to 0, \quad n \to \infty. \qquad \square$$

Lemma 3.2.4. *Under Assumption* (3.2.7), *with* $L^{\mathrm{J}}(t)$ *and* $L^{\mathrm{J}}(n,t)$ *as defined in* (3.2.8) *and* (3.2.9), *as* $n \to \infty$, *we have*

$$\rho\left(L^{\mathrm{J}}(n,t), L^{\mathrm{J}}(t)\right) \xrightarrow{\mathrm{P}} 0.$$

Proof. Take $k \in \{1, \ldots, N(n)\}$ and $\tau_k(n)$ as given by (3.2.11). For a given path ω, define $\lambda(n,t;\omega)$ for $t \in G(n)$ as

$$\lambda(n, k\Delta t(n); \omega) = \begin{cases} 0, & k = 1, \\ T, & k = N(n), \\ \tau_k(n;\omega) \wedge k\Delta t(n), & \text{otherwise}, \end{cases}$$

and interpolate piecewise linearly to obtain $\lambda(n,t;\omega)$ for $t \in [0,T] \setminus G(n)$. Note that $\lambda(n,\cdot;\omega) \in \Lambda$ and

$$\sup_{t\in[0,T]} |\lambda(n,t;\omega) - t| \leq \Delta t(n) \to 0, \quad n \to \infty.$$

The time change $\lambda(n,t)$ is similar to that used in the proof for Theorem 3.2(i) in Szimayer and Maller (2007). Put simply, it acts to align the jumps of $L^{(3,2)}(n,t)$ with those of $L^{\mathrm{PJ}}(n,t)$, which fall on $G(n)$, in all but the last interval.

Now, by Lemmas 3.2.1 and 3.2.2,

$$\sup_{t\in[0,T]} |L^{\mathrm{J}}(n,t) - L^{\mathrm{J}}(\lambda(n,t))| \leq T\,\|\gamma^{\mathrm{J}}(n) - \gamma(n)\| + \|\Delta L^{\mathrm{PJ}}(n,T)\| + \Delta t(n) \max\left(\|\gamma^{\mathrm{J}}(n)\|, \|\gamma(n)\|\right) + R(n),$$

where $R(n) \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$. The first term on the right hand side also vanishes as $n \to \infty$ by Lemma 3.2.3.

By the definition of $L^{\mathrm{PJ}}(n,t)$ in terms of the sequence of random variables $\{X(n,i)\}_{i\in\{1,\ldots,N(n)\}}$ in (3.2.3), $\Delta L^{\mathrm{PJ}}(n,T) = X(n,N(n))$. Equation (3.2.4) gives that $\mathrm{P}(X(n,\cdot) \neq 0) = 1 - \mathrm{e}^{-\Delta t(n)\Pi(\mathcal{J}(n))}$, which tends to zero by (3.2.15), so $\|\Delta L^{\mathrm{PJ}}(n,T)\| \xrightarrow{\mathrm{P}} 0$. Also note that by applying (3.2.16) we obtain

$$\Delta t(n)\gamma_j(n) = \Delta t(n)\left(r - \frac{\sigma_j^2}{2}\right) - \Delta t(n)\Pi(\mathcal{J}(n))\left\{\int_{\mathcal{J}(n)} \mathrm{e}^{x_j}\frac{\Pi(dx)}{\Pi(\mathcal{J}(n)}) - 1\right\} - \mathrm{O}\left((\Delta t(n)\Pi(\mathcal{J}(n)))^2\right), \quad j \in \{1,\ldots,d\} \tag{3.2.17}$$

which in conjunction with Lemma 3.2.3 gives $\Delta t(n)\max\left(\|\gamma^{\mathrm{J}}(n)\|, \|\gamma(n)\|\right) \to 0$ as $n \to \infty$.

We have shown that $\rho\left(L^{\mathrm{J}}(n,t), L^{\mathrm{J}}(t)\right) \xrightarrow{\mathrm{P}} 0$. □

The following proof is similar to that for Theorem 3.3(i) in Szimayer and Maller (2007), but is much simplified as we do not require our first jump approximation to live on a discrete grid.

Proof of Theorem 3.2.1(i). Recall that $B(t)$ is the Gaussian part of $L(t)$ and

$$B(n,t) = B\left(\left\lfloor \frac{t}{\Delta t(n)} \right\rfloor \Delta t(n)\right), \quad t \in [0,T].$$

The Lévy modulus of continuity (Karatzas and Shreve 1991, p. 114) therefore gives that for $j \in \{1,\ldots,d\}$, $k \in \{0,\ldots,N(n)-1\}$, and $\Delta t(n) < 1$,

$$\lim_{n\to\infty} \sup_{s\in[0,\Delta t(n)]} \frac{|B_j(k\Delta t(n)+s) - B_j(n,k\Delta t(n)+s)|}{\sigma_j\sqrt{2\Delta t(n)\log\left(\Delta t(n)^{-1}\right)}} = 1 \quad \text{a.s.}$$

This implies that $\sup_{t\in[0,T]}\|B(t) - B(t,n)\| \xrightarrow{\text{a.s.}} 0$ so $\rho(B(n,t),B(t)) \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$. Noting that $L(t) = L^{\mathrm{J}}(t) + B(t)$ and $L(n,t) = L^{\mathrm{J}}(n,t) + B(n,t)$, and that $B(t)$ is a continuous process, Proposition 1.23, Chapter VI of Jacod and Shiryaev

(2002, p. 329) in combination with Lemma 3.2.4 gives that $\rho(L(n,t), L(t)) \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$. □

Lemma 3.2.5. *Under Assumption* (3.2.7)*, the sequence of processes* $L^{\mathrm{J}}(n,t)$ *defined in* (3.2.9) *satisfies*

$$\lim_{\delta \searrow 0} \limsup_{n \to \infty} \sup_{\sigma,\tau \in \mathcal{S}^{\mathrm{D}}_{0,T}(n); \sigma \leq \tau \leq \sigma + \delta} \mathrm{P}(\|L^{\mathrm{J}}(n,\tau) - L^{\mathrm{J}}(n,\sigma)\| \geq \varepsilon) = 0,$$

where $\mathcal{S}^{D}_{0,T}(n)$ *is the set of* $\mathbb{F}^{L^D}(n)$*-stopping times in* $[0,T]$ *and* $\mathbb{F}^{L^D}(n) = \{\mathcal{F}_t^{L^D}(n)\}_{t \geq 0}$ *is the filtration generated by* $L^{D}(n,t)$*, which is defined in* (3.3.1)*.*

Remark 3.2.2. *The above expression is very similar to Aldous' criterion for tightness in probability, except that the set of stopping times is a superset of the* $\mathbb{F}^{L^J(n,t)}$*-stopping times on* $[0,T]$*. It is the set of stopping times measurable on the filtration generated by the skeleton of* $L(t)$ *on* $G(n)$*, the time grid. This more general set is required to make this lemma useful to later theorems.*

Lemma 3.2.5 is a generalisation of Theorem 3.2(ii) from Szimayer and Maller (2007). Lemma 3.2.5 differs by using a different approximation scheme, a wider set of stopping times, and arbitrary dimensions. The respective proofs develop with subtle differences.

Proof. Take $\varepsilon, \delta > 0$ and $\sigma, \tau \in \mathcal{S}^{\mathrm{D}}_{0,T}(n)$ with $\tau \in [\sigma, \sigma + \delta]$. Using $L^{\mathrm{S}}(n,t)$ as defined in (3.2.14),

$$\begin{aligned} \mathrm{P}\left(\|L^{\mathrm{J}}(n,\tau) - L^{\mathrm{J}}(n,\sigma)\| \geq \varepsilon\right) \leq{}& \mathrm{P}\left(\|L^{\mathrm{S}}(n,\tau) - L^{\mathrm{S}}(n,\sigma)\| \geq \frac{\varepsilon}{3}\right) \\ &+ \mathrm{P}\left(\|L^{\mathrm{J}}(n,\tau) - L^{\mathrm{S}}(n,\tau)\| \geq \frac{\varepsilon}{3}\right) \\ &+ \mathrm{P}\left(\|L^{\mathrm{J}}(n,\sigma) - L^{\mathrm{S}}(n,\sigma)\| \geq \frac{\varepsilon}{3}\right). \end{aligned}$$

The following reasoning holds equally well for σ and τ. Define

$$t_\sigma(n) = \left\lfloor \frac{\sigma}{\Delta t(n)} \right\rfloor \Delta t(n)$$

and recall the relationship between $L^{\mathrm{PJ}}(n,t)$ and $X(n,i)$ for $i \in \{1,\dots,N(n)\}$ given in (3.2.3) to see that

$$\begin{aligned}\|L^{\mathrm{J}}(n,\sigma) - L^{\mathrm{S}}(n,\sigma)\| &\le \|L^{\mathrm{J}}(n,t_\sigma(n)+\Delta t(n)) - L^{\mathrm{J}}(n,t_\sigma(n))\| \\ &\le \left\| X\left(n, \frac{t_\sigma(n)}{\Delta t(n)} + 1\right)\right\| + \|\gamma(n) - \gamma^{\mathrm{J}}(n)\|\,\Delta t(n).\end{aligned}$$

We know from Lemma 3.2.3 that $\|\gamma(n) - \gamma^{\mathrm{J}}(n)\| \to 0$ as $n \to \infty$, allowing us to assume that for sufficiently large n, $\|\gamma(n) - \gamma^{\mathrm{J}}(n)\|\,\Delta t(n) < \varepsilon/3$, so

$$\begin{aligned}&\mathrm{P}\left(\|L^{\mathrm{J}}(n,\sigma) - L^{\mathrm{S}}(n,\sigma)\| \ge \frac{\varepsilon}{3}\right) \\ &\le \mathrm{P}\left(\left\| X\left(n, \frac{t_\sigma(n)}{\Delta t(n)} + 1\right)\right\| \ge \frac{\varepsilon}{3} - \|\gamma(n) - \gamma^{\mathrm{J}}(n)\|\,\Delta t(n)\right) \\ &\le \mathrm{P}\left(\left\| X\left(n, \frac{t_\sigma(n)}{\Delta t(n)} + 1\right)\right\| > 0\right).\end{aligned}$$

For such an n, as σ is $\mathcal{F}^{L^{\mathrm{D}}}_{t_\sigma(n)}(n)$-measurable and $X(n, t_\sigma(n)/\Delta t(n)+1)$ is independent of $\mathcal{F}^{L^{\mathrm{D}}}_{t_\sigma(n)}(n)$,

$$\begin{aligned}&\mathrm{P}\left(\|L^{\mathrm{J}}(n,\tau) - L^{\mathrm{S}}(n,\tau)\| \ge \frac{\varepsilon}{3}\right) + \mathrm{P}\left(\|L^{\mathrm{J}}(n,\sigma) - L^{\mathrm{S}}(n,\sigma)\| \ge \frac{\varepsilon}{3}\right) \\ &\le 2\,\mathrm{P}\left(\|X(n,1)\| > 0\right) = 2\left(1 - \mathrm{e}^{-\Delta t(n)\Pi(\mathcal{J}(n))}\right) \to 0, \quad n \to \infty, \end{aligned} \tag{3.2.18}$$

where the equality follows from the distribution of $X(n,1)$ given in (3.2.4). Turning

now to the remaining probability term,

$$\begin{aligned}
&\mathrm{P}\left(\|L^{\mathrm{S}}(n,\tau) - L^{\mathrm{S}}(n,\sigma)\| \geq \frac{\varepsilon}{3}\right) \\
&\leq \mathrm{P}\left(\|L^{\mathrm{J}}(\tau) - L^{\mathrm{J}}(\sigma)\| \geq \frac{\varepsilon}{6}\right) \\
&\quad + \mathrm{P}\left(\|L^{\mathrm{S}}(n,\sigma) - L^{\mathrm{J}}(\sigma)\| + \|L^{\mathrm{S}}(n,\tau) - L^{\mathrm{J}}(\tau)\| \geq \frac{\varepsilon}{6}\right).
\end{aligned}$$

Attending to the second term on the right hand side first,

$$\|L^{\mathrm{S}}(n,\sigma) - L^{\mathrm{J}}(\sigma)\| + \|L^{\mathrm{S}}(n,\tau) - L^{\mathrm{J}}(\tau)\| \leq 2 \sup_{t\in[0,T]} \|L^{\mathrm{S}}(n,t) - L^{\mathrm{J}}(t)\|,$$

and by equations (3.2.10), (3.2.12), and (3.2.14), and the triangle inequality,

$$\|L^{\mathrm{S}}(n,t) - L^{\mathrm{J}}(t)\| \leq \left\|L^{(1)}(n,t)\right\| + \left\|L^{(2)}(n,t)\right\| + \left\|L^{(3,1)}(n,t)\right\|,$$

so

$$\begin{aligned}
&\mathrm{P}\left(\|L^{\mathrm{S}}(n,\sigma) - L^{\mathrm{J}}(\sigma)\| + \|L^{\mathrm{S}}(n,\tau) - L^{\mathrm{J}}(\tau)\| \geq \frac{\varepsilon}{6}\right) \\
&\leq \mathrm{P}\left(\sup_{t\in[0,T]} \left\|L^{(1)}(n,t)\right\| \geq \frac{\varepsilon}{36}\right) + \mathrm{P}\left(\sup_{t\in[0,T]} \left\|L^{(2)}(n,t)\right\| \geq \frac{\varepsilon}{36}\right) \\
&\quad + \mathrm{P}\left(\sup_{t\in[0,T]} \left\|L^{(3,1)}(n,t)\right\| \geq \frac{\varepsilon}{36}\right) \to 0, \quad n \to \infty,
\end{aligned}$$

by Lemmas 3.2.1 and 3.2.2.

Observing that $\mathrm{P}\left(\|L^{\mathrm{J}}(\tau) - L^{\mathrm{J}}(\sigma)\| \geq \frac{\varepsilon}{6}\right) \leq \sum_{j=1}^{d} \mathrm{P}\left(\left|L_j^{\mathrm{J}}(\tau) - L_j^{\mathrm{J}}(\sigma)\right| \geq \frac{\varepsilon}{6d}\right)$, we choose a $j \in \{1,\ldots,d\}$. As $\mathbb{F}^{L^{\mathrm{D}}}$ is the filtration generated by a skeleton of $L(t)$, we have that $\mathbb{F}^{L^{\mathrm{D}}} \subset \mathbb{F}^{L}$. Thus σ is also an $\mathbb{F}^{L}$-stopping time, so by the strong Markov property and stationarity of $L^{\mathrm{J}}(t)$, the jump part of $L(t)$,

$$\mathrm{P}\left(\left|L_j^{\mathrm{J}}(\tau) - L_j^{\mathrm{J}}(\sigma)\right| \geq \frac{\varepsilon}{6d}\right) \leq \mathrm{P}\left(\sup_{t\in[0,\delta]} \left|L_j^{\mathrm{J}}(t)\right| \geq \frac{\varepsilon}{6d}\right). \tag{3.2.19}$$

We now decompose $L_j^{\mathrm{J}}(t)$ as $L_j^{\mathrm{J}}(t) = \gamma t + L_j^{\mathrm{J}(1)}(t) + L_j^{\mathrm{J}(2)}(t)$, where $L_j^{\mathrm{J}(1)}(t)$ is the small jump martingale capturing jumps in $[-1, 1]$ and $L_j^{\mathrm{J}(2)}(t)$ accounts for jumps outside $[-1, 1]$. If we further restrict $\delta < \varepsilon/12d$, we obtain

$$\begin{aligned}
&\mathrm{P}\left(\sup_{t\in[0,\delta]} \left|L_j^{\mathrm{J}}(t)\right| \geq \frac{\varepsilon}{6d}\right) \\
&\leq \mathrm{P}\left(\sup_{t\in[0,\delta]} \left|L_j^{\mathrm{J}(1)}(t)\right| \geq \frac{\varepsilon}{12d}\right) + \mathrm{P}\left(\sup_{t\in[0,\delta]} \left|L_j^{\mathrm{J}(2)}(t)\right| > 0\right).
\end{aligned}$$

Chebyshev's inequality and Doob's maximal quadratic inequality give

$$\begin{aligned}
\mathrm{P}\left(\sup_{t\in[0,\delta]} \left|L_j^{\mathrm{J}(1)}(t)\right| \geq \frac{\varepsilon}{12d}\right) &\leq \left(\frac{12d}{\varepsilon}\right)^2 \mathrm{E}\left(\sup_{t\in[0,\delta]} \left|L_j^{\mathrm{J}(1)}(t)\right|\right)^2 \\
&\leq \left(\frac{24d}{\varepsilon}\right)^2 \mathrm{E}\left|L_j^{\mathrm{J}(1)}(\delta)\right|^2.
\end{aligned}$$

Example 25.12 in Sato (1999, p. 163) then shows that

$$\mathrm{E}\left|L_j^{\mathrm{J}(1)}(\delta)\right|^2 = \delta \int_{|x_j|\leq 1} x_j^2 \Pi(\mathrm{d}x).$$

Sato (1999, p. 136) also allows us to write

$$\mathrm{P}\left(\sup_{t\in[0,\delta]} \left|L_j^{\mathrm{J}(2)}(t)\right| > 0\right) = 1 - \mathrm{e}^{-\delta\Pi(\{|x_j|>1\})},$$

which is the probability that no jumps of magnitude greater than one occur in the time period.

We therefore have that

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{\sigma,t\in\mathcal{S}_{0,T}^{\mathrm{D}}(n),\sigma\leq\tau\leq\sigma+\delta} \mathrm{P}(\|L^{\mathrm{J}}(n,\tau) - L^{\mathrm{J}}(n,\sigma)\| \geq \varepsilon) \\
&\leq \sum_{j=1}^{d} \left\{1 - \mathrm{e}^{-\delta\Pi(\{|x_j|>1\})} + \delta\left(\frac{24d}{\varepsilon}\right)^2 \int_{|x_j|\leq 1} x_j^2 \Pi(\mathrm{d}x)\right\},
\end{aligned}$$

which tends to zero as $\delta \to 0$. □

Theorem 3.2.1(iii) is a generalisation of Theorem 3.3(ii) from Szimayer and Maller (2007). The proofs are similar but differ at some points, as we use a different approximation for the Gaussian part of $L(t)$.

Proof of Theorem 3.2.1(iii). Let $\varepsilon, \delta > 0$ and $\sigma, \tau \in \mathcal{S}^{\mathrm{D}}_{0,T}(n)$ with $\tau \in [\sigma, \sigma + \delta]$. $\mathcal{S}^{\mathrm{D}}_{0,T}(n)$ is the set of $\mathbb{F}^{L^{\mathrm{D}}}(n)$-stopping times in $[0,T]$ and $\mathbb{F}^{L^{\mathrm{D}}}(n) = \{\mathcal{F}_t^{L^{\mathrm{D}}}(n)\}_{t\geq 0}$ is the filtration generated by $L^{\mathrm{D}}(n,t)$, the skeleton of $L(t)$ on $G(n)$, as defined in (3.3.1). Using (3.2.1) and (3.2.9) write

$$\begin{aligned}
&\mathrm{P}(\|L(n,\tau) - L(n,\sigma)\| \geq \varepsilon) \\
&\leq \mathrm{P}\left(\|L^{\mathrm{J}}(n,\tau) - L^{\mathrm{J}}(n,\sigma)\| \geq \frac{\varepsilon}{2}\right) + \mathrm{P}\left(\|B(n,\tau) - B(n,\sigma)\| \geq \frac{\varepsilon}{2}\right).
\end{aligned}$$

Lemma 3.2.5 shows that the first term on the right hand side tends to zero as $n \to \infty$ and $\delta \to 0$. For the Gaussian term, we have that

$$\begin{aligned}
&\mathrm{P}\left(\|B(n,\tau) - B(n,\sigma)\| \geq \frac{\varepsilon}{2}\right) \\
&\leq \mathrm{P}\left(\|B(\tau) - B(\sigma)\| \geq \frac{\varepsilon}{6}\right) + \mathrm{P}\left(\|B(n,\tau) - B(\tau)\| \geq \frac{\varepsilon}{6}\right) \\
&\quad + \mathrm{P}\left(\|B(n,\sigma) - B(\sigma)\| \geq \frac{\varepsilon}{6}\right).
\end{aligned}$$

The first term on the right hand side does not depend on n and tends to zero as $\delta \to 0$ because $B(t)$ is stochastically continuous.

Note that

$$\mathrm{P}\left(\|B(n,\sigma) - B(\sigma)\| \geq \frac{\varepsilon}{2}\right) \leq \sum_{j=1}^{d} \mathrm{P}\left(|B_j(n,\sigma) - B_j(\sigma)| \geq \frac{\varepsilon}{2d}\right)$$

and choose $j \in \{1, \ldots, d\}$. Take $t_\sigma(n) = \lfloor \sigma/\Delta t(n) \rfloor \Delta t(n)$. By definition, $B_j(\sigma) - B_j(n,\sigma) = B_j(\sigma) - B_j(t_\sigma(n))$, but $t_\sigma(n)$ is $\mathcal{F}^{L^{\mathrm{D}}}_{t_\sigma(n)}(n)$-measurable and by construction $\mathcal{F}^{L^{\mathrm{D}}}_{t_\sigma(n)}(n) \subset \mathcal{F}^{L}_{t_\sigma(n)}$, so $t_\sigma(n)$ is also a $\mathbb{F}^L$-stopping time. This allows us to use the

strong Markov property and stationarity of $B_j(t)$ to write

$$\mathrm{P}\left(|B_j(n,\tau) - B_j(\tau)| \geq \frac{\varepsilon}{6d}\right) \leq \mathrm{P}\left(\sup_{t\in[0,\Delta t(n)]} |B_j(t)| \geq \frac{\varepsilon}{6d}\right).$$

As $B(t)$ is a martingale we can apply Chebyshev's inequality and Doob's maximal quadratic inequality to obtain

$$\mathrm{P}\left(\sup_{t\in[0,\Delta t(n)]} |B_j(t)| \geq \frac{\varepsilon}{6d}\right) \leq \left(\frac{12d}{\varepsilon}\right)^2 \mathrm{E}\, B^2(\Delta t(n)) = \left(\frac{12d}{\varepsilon}\right)^2 \sigma_j^2 \Delta t(n),$$

where σ_j^2 is the jth diagonal element of the Gaussian covariance matrix A. Similar reasoning for τ allows us to write

$$\mathrm{P}\left(|B_j(n,\tau) - B_j(\tau)| \geq \frac{\varepsilon}{6d}\right) + \mathrm{P}\left(|B_j(n,\sigma) - B_j(\sigma)| \geq \frac{\varepsilon}{6d}\right) \to 0,$$

for $n \to \infty$.

As $\mathcal{S}_{0,T}(n) \subset \mathcal{S}^{\mathrm{D}}_{0,T}(n)$,

$$\begin{aligned} &\sup_{\sigma,t\in\mathcal{S}_{0,T}(n);\sigma\leq\tau\leq\sigma+\delta} \mathrm{P}(\|L^{\mathrm{J}}(n,\tau) - L^{\mathrm{J}}(n,\sigma)\| \geq \varepsilon) \\ &\leq \sup_{\sigma,t\in\mathcal{S}^{\mathrm{D}}_{0,T}(n);\sigma\leq\tau\leq\sigma+\delta} \mathrm{P}(\|L^{\mathrm{J}}(n,\tau) - L^{\mathrm{J}}(n,\sigma)\| \geq \varepsilon), \end{aligned}$$

but we have shown that the right hand side tends to zero as $n \to \infty$ and $\delta \to 0$, so we have the required result. □

Remark 3.2.3. *As for Lemma 3.2.5, we have shown the above result for a slightly larger set of stopping times than was immediately necessary. This will be useful later.*

The next two lemmas are generalisations of Lemma 3 and Proposition 2 in Coquet, Mémin, and Słominski (2001). Their proofs are very similar to those in that paper.

Lemma 3.2.6. *Define the random process* $X(\cdot) \in \mathbb{D}\left([0,T], \mathbb{R}^d\right)$ *with natural filtration* $(\mathcal{F}_t)_{t\in[0,T]} = \mathbb{F}$. *Let* $(\mathcal{F}_t(n))_{t\in[0,T]} = \mathbb{F}(n)$, $n \in \mathbb{N}$, *be a sequence of filtrations. As* $n \to \infty$, *the following three conditions are equivalent:*

(i) $\mathbb{F}(n) \xrightarrow{\text{W}} \mathbb{F}$,

(ii) for all bounded and continuous $f : \mathbb{R}^{d\times k} \mapsto \mathbb{R}$ *and all points of continuity of* $X(t)$, t_ℓ, $\ell \in \{1,\ldots,k\}$, $k \in \mathbb{N}$,

$$\rho(\mathrm{E}(f(X(t_1),\ldots,X(t_k))|\mathcal{F}_\cdot(n)), \mathrm{E}(f(X(t_1),\ldots,X(t_k))|\mathcal{F}_\cdot)) \xrightarrow{\text{P}} 0, \quad \textit{and}$$

(iii) for all $a \in \mathbb{R}^m$, $b \in \mathbb{R}^{m\times k\times d}$, *and all points of continuity of* $X(t)$, t_ℓ, $\ell \in \{1,\ldots,k\}$, $k, m \in \mathbb{N}$,

$$\rho\left(\mathrm{E}\left(\sum_{p=1}^{m} a_p \exp\left\{\mathrm{i}\sum_{\ell=1}^{k}\sum_{j=1}^{d} b_{p,\ell,j} X_j(t_\ell)\right\}\middle| \mathcal{F}_\cdot(n)\right),\right.$$
$$\left.\mathrm{E}\left(\sum_{p=1}^{m} a_p \exp\left\{\mathrm{i}\sum_{\ell=1}^{k}\sum_{j=1}^{d} b_{p,\ell,j} X_j(t_\ell)\right\}\middle| \mathcal{F}_\cdot\right)\right) \xrightarrow{\text{P}} 0.$$

Proof. Conditions (ii) and (iii) follow from condition (i) by Lemma 2.5.1.

To see that condition (i) is implied by condition (ii), assume condition (ii) holds. Take a sequence $\lambda(n,t) \in \Lambda$. For any $A \in \mathcal{F}_T$,

$$\begin{aligned}
&\sup_{t\in[0,T]} \left|\mathrm{E}\left(\mathbf{1}_A \,|\mathcal{F}_t(n)\right) - \mathrm{E}\left(\mathbf{1}_A \,\middle|\mathcal{F}_{\lambda(n,t)}\right)\right| \\
&\leq \sup_{t\in[0,T]} \left|\mathrm{E}\left(f(X(t_1),\ldots,X(t_k)) \,|\mathcal{F}_t(n)\right) - \mathrm{E}\left(f(X(t_1),\ldots,X(t_k)) \,\middle|\mathcal{F}_{\lambda(n,t)}\right)\right| \\
&\quad + \sup_{t\in[0,T]} \left|\mathrm{E}\left(\mathbf{1}_A - f(X(t_1),\ldots,X(t_k)) \,|\mathcal{F}_t(n)\right)\right| \\
&\quad + \sup_{t\in[0,T]} \left|\mathrm{E}\left(\mathbf{1}_A - f(X(t_1),\ldots,X(t_k)) \,|\mathcal{F}_t\right)\right|,
\end{aligned}$$

where $f : \mathbb{R}^{d\times k} \mapsto \mathbb{R}$ is bounded and continuous and t_ℓ, $\ell \in \{1,\ldots,k\}$, $k \in \mathbb{N}$, are points of continuity of $X(t)$. By condition (ii), we can find a sequence $\lambda(n,t)$ such

that $\sup_{t\in[0,T]} |\lambda(n,t) - t| \xrightarrow{\mathrm{P}} 0$ and

$$\sup_{t\in[0,T]} \left| \mathrm{E}\left(f(X(t_1),\ldots,X(t_k))\,|\mathcal{F}_t(n)\right) - \mathrm{E}\left(f(X(t_1),\ldots,X(t_k))\,\middle|\mathcal{F}_{\lambda(n,t)}\right)\right| \xrightarrow{\mathrm{P}} 0.$$

Doob's maximal inequality for martingales gives that for $\varepsilon > 0$,

$$\begin{aligned}
&\mathrm{P}\left(\sup_{t\in[0,T]} |\mathrm{E}\left(\mathbf{1}_A - f(X(t_1),\ldots,X(t_k))\,|\mathcal{F}_t(n)\right)| > \varepsilon\right) \\
&\leq \frac{1}{\varepsilon}\,\mathrm{E}\,|\,\mathbf{1}_A - f(X(t_1),\ldots,X(t_k))| \quad \text{and} \\
&\mathrm{P}\left(\sup_{t\in[0,T]} |\mathrm{E}\left(\mathbf{1}_A - f(X(t_1),\ldots,X(t_k))\,|\mathcal{F}_t\right)| > \varepsilon\right) \\
&\leq \frac{1}{\varepsilon}\,\mathrm{E}\,|\,\mathbf{1}_A - f(X(t_1),\ldots,X(t_k))|.
\end{aligned}$$

As in the proof for Lemma 3 in Coquet et al. (2001), for $\varepsilon' > 0$ we can always choose bounded and continuous f and continuity points $\{t_\ell\}_{\ell\in\{1,\ldots,k\}}$ such that $\mathrm{E}\,|\,\mathbf{1}_A - f(X(t_1),\ldots,X(t_k))| < \varepsilon'$. We therefore have that condition (ii) implies condition (i).

Proof that condition (ii) follows from condition (iii) proceeds in a similar fashion, differing only in that we require instead that for any $k \in \mathbb{N}$, bounded and continuous f, continuity points $\{t_\ell\}_{\ell\in\{1,\ldots,k\}}$, and $\varepsilon' > 0$, there exist $m \in \mathbb{N}$, $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^{m\times k\times d}$ such that

$$\mathrm{E}\left|f(X(t_1),\ldots,X(t_k)) - \sum_{p=1}^{m} a_p \exp\left\{\mathrm{i}\sum_{\ell=1}^{k}\sum_{j=1}^{d} b_{p,\ell,j} X_j(t_\ell)\right\}\right| < \varepsilon',$$

but such m, a, and b can always be found, so the result follows. □

Lemma 3.2.7. *Let $X(t)$ and $\{X(n,t)\}_{n\in\mathbb{N}}$ be random processes drawn from $\mathbb{D}\left([0,T],\mathbb{R}^d\right)$ with natural filtrations $\left(\mathcal{F}_t^X\right)_{t\in[0,T]} = \mathbb{F}^X$ and $\left(\mathcal{F}_t^X(n)\right)_{t\in[0,T]} = \mathbb{F}^X(n)$. If $\{X(n,t)\}_{n\in\mathbb{N}}$ have independent increments and $\rho(X(n,\cdot),X(\cdot)) \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$, then $\mathbb{F}^X(n) \xrightarrow{\mathrm{W}} \mathbb{F}^X$.*

Proof. We will show that condition (iii) of Lemma 3.2.6 is satisfied. Take $\{t_\ell\}_{\ell\in\{1,\ldots,k\}}$, $k \in \mathbb{N}$, to be an increasing set of continuity points of $X(t)$ and $m \in \mathbb{N}$, $a \in \mathbb{R}^m$, and $b \in \mathbb{R}^{m\times k\times d}$. Reparametrise b as $b'_{p,\ell,j} = \sum_{u=\ell}^{k} b_{p,u,j}$ so that

$$\begin{aligned}
&\sum_{p=1}^{m} a_p \exp\left\{ \mathrm{i} \sum_{\ell=1}^{k} \sum_{j=1}^{d} b_{p,\ell,j} X_j(t_\ell) \right\} \\
&= \sum_{p=1}^{m} a_p \exp\left\{ \mathrm{i} \sum_{\ell=1}^{k} \sum_{j=1}^{d} b'_{p,\ell,j} [X_j(t_\ell) - X_j(t_{\ell-1})] \right\},
\end{aligned}$$

where we have set $t_0 < 0$ and $X(t_0) = 0$ for notational convenience. As finite diminsional distributions of $X(n,\cdot)$ converge in law to the corresponding finite dimensional distributions of $X(\cdot)$ except on countable subsets of $[0,T]$, see Billingsley (1999, p. 138), $X(t)$ also has independent increments. Define $u(t) = \max\{\ell \in \{0,\ldots,k\} : t_\ell \le t\}$ and the processes

$$\begin{aligned}
F(t) &= \mathrm{E}\left(\sum_{p=1}^{m} a_p \exp\left\{ \mathrm{i} \sum_{\ell=1}^{k} \sum_{j=1}^{d} b'_{p,\ell,j} [X_j(t_\ell) - X_j(t_{\ell-1})] \right\} \middle| \mathcal{F}_t^X \right) \\
&= \sum_{p=1}^{m} a_p \prod_{\ell=1}^{u(t)} \exp\left\{ \mathrm{i} \sum_{j=1}^{d} b'_{p,\ell,j} [X_j(t_\ell) - X_j(t_{\ell-1})] \right\} \\
&\quad \times \exp\left\{ \mathrm{i} \sum_{j=1}^{d} b'_{p,u(t),j} \left[X_j(t) - X_j\left(t_{u(t)}\right) \right] \right\} \\
&\quad \times \mathrm{E} \exp\left\{ \mathrm{i} \sum_{j=1}^{d} b'_{p,u(t),j} \left[X_j\left(t_{u(t)+1}\right) - X_j(t) \right] \right\} \\
&\quad \times \prod_{\ell=u(t)+1}^{k} \mathrm{E} \exp\left\{ \mathrm{i} \sum_{j=1}^{d} b'_{p,\ell,j} [X_j(t_\ell) - X_j(t_{\ell-1})] \right\}
\end{aligned}$$

and

$$
\begin{aligned}
F(n,t) &= \mathrm{E}\left(\sum_{p=1}^{m} a_p \exp\left\{ \mathrm{i} \sum_{\ell=1}^{k} \sum_{j=1}^{d} b'_{p,\ell,j} [X_j(n,t_\ell) - X(t_{\ell-1})] \right\} \middle| \mathcal{F}_t^X(n) \right) \\
&= \sum_{p=1}^{m} a_p \prod_{\ell=1}^{u(t)} \exp\left\{ \mathrm{i} \sum_{j=1}^{d} b'_{p,\ell,j} [X_j(n,t_\ell) - X_j(n,t_{\ell-1})] \right\} \\
&\quad \times \exp\left\{ \mathrm{i} \sum_{j=1}^{d} b'_{p,u(t),j} \left[X_j(n,t) - X_j\left(n, t_{u(t)}\right) \right] \right\} \\
&\quad \times \mathrm{E} \exp\left\{ \mathrm{i} \sum_{j=1}^{d} b'_{p,u(t),j} \left[X_j\left(n, t_{u(t)+1}\right) - X_j(n,t) \right] \right\} \\
&\quad \times \prod_{\ell=u(t)+1}^{k} \mathrm{E} \exp\left\{ \mathrm{i} \sum_{j=1}^{d} b'_{p,\ell,j} [X_j(n,t_\ell) - X_j(n,t_{\ell-1})] \right\}.
\end{aligned}
$$

As $\{t_\ell\}_{\ell \in \{1,\ldots,k\}}$ are points of continuity of $X(t)$, $\rho(X(n,\cdot), X(\cdot)) \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$ implies that $\rho(F(n,\cdot), F(\cdot)) \xrightarrow{\mathrm{P}} 0$. It then follows from Lemma 3.2.6 that $\mathbb{F}^X(n) \xrightarrow{\mathrm{W}} \mathbb{F}^X$. □

The following proof completes that of Theorem 3.2.1.

Proof of Theorem 3.2.1(ii). Note that $L(n,t)$ has independent increments. Theorem 3.2.1(ii) then follows from Theorem 3.2.1(i) and Lemma 3.2.7. This concludes the proof of Theorem 3.2.1. □

3.3 The Skeleton Approximation Model

The model formulated in the previous section can be constructed for any Lévy process for which the Lévy measure of $L(t)$ can be calculated. The following model is useful for Lévy processes where the transition probability density function of $L(t)$ can be efficiently evaluated.

Take $\Delta t(n)$ as defined in Section 3.2. We define the skeleton process

$$L^{\mathrm{D}}(n,t) = L\left(\left\lfloor \frac{t}{\Delta t(n)} \right\rfloor \Delta t(n)\right), \quad t \in [0,T], \tag{3.3.1}$$

the corresponding stock price process

$$S^{\mathrm{D}}(n,t) = \left(S_1(0)\mathrm{e}^{L_1^{\mathrm{D}}(n,t)}, \ldots, S_d(0)\mathrm{e}^{L_d^{\mathrm{D}}(n,t)}\right), \tag{3.3.2}$$

and, outside the exercise region, the option price process

$$\pi^{\mathrm{D}}(n,t) = \operatorname*{ess\,sup}_{\tau \in \mathcal{S}_{t,T}^{\mathrm{D}}(n)} \mathrm{E}\left(\mathrm{e}^{-r(\tau-t)} h(\tau, S^{\mathrm{D}}(n,\tau)) \middle| \mathcal{F}_t^{L^{\mathrm{D}}}(n)\right), \tag{3.3.3}$$

where $\mathcal{S}_{t,T}^{\mathrm{D}}(n)$ is the set of $\mathbb{F}^{L^{\mathrm{D}}}(n)$-stopping times in $[t,T]$ and $\mathbb{F}^{L^{\mathrm{D}}}(n) = \{\mathcal{F}_t^{L^{\mathrm{D}}}(n)\}_{t\geq 0}$ is the filtration induced on Ω by $L^{\mathrm{D}}(n,t)$.

$L^{D}(n,t)$ converges to $L(t)$ as follows:

Theorem 3.3.1. *As $n \to \infty$,*

(i) $\rho\left(L^{D}(n,\cdot), L(\cdot)\right) \xrightarrow{\mathrm{P}} 0$,

(ii) $\mathbb{F}^{L^{D}}(n) \xrightarrow{\mathrm{W}} \mathbb{F}^{L}$, *and*

(iii) $\{L^{D}(n,t)\}_{n\in\mathbb{N}}$ *fulfills Aldous' criterion for tightness in probability.*

This result, in one or d dimensions, has not been considered in the literature before. All the proofs in this section are new.

Theorem 3.3.1 can be proved using the developments in Section 3.2. For $n \in \mathbb{N}$, take $G(n)$, $\mathcal{R}(\cdot)$, $m(n)$, and the Lévy measure Π as defined in Sections 3.1 and 3.2. Note that $\mathcal{R}(\cdot)$ and $m(n)$ are not a part of the skeleton approximation model specification. They are used here to specify a first jump approximation to $L(t)$, which is then used in the proof of Theorem 3.3.1. Construct them to have the

property

$$\lim_{n\to\infty} \Delta t(n)\Pi^2\left(\mathbb{R}^d \setminus \mathcal{R}(m(n))\right) = 0. \tag{3.3.4}$$

Take the resulting first jump approximation $L(n,t)$, as given in (3.2.1), and define the remainder process

$$L^{\mathrm{R}}(n,t) = L^{\mathrm{D}}(n,t) - L(n,t), \quad t \in [0,T]. \tag{3.3.5}$$

Lemma 3.3.1. *For $L^{\mathrm{R}}(n,t)$ defined in (3.3.5), as $n \to \infty$,*

$$\sup_{t\in[0,T]} \|L^{\mathrm{R}}(n,t)\| \xrightarrow{\mathrm{P}} 0.$$

Proof. Using (3.2.1), (3.3.1), (3.1.1), (3.2.8), (3.2.10), (3.2.12), and (3.2.13) and the terms defined therein, for $t \in G(n)$,

$$\begin{aligned} L(n,t) &= \gamma(n)t + B(t) + L^{(3,2)}(n,t) \quad \text{and} \\ L^{\mathrm{D}}(n,t) &= \gamma^{\mathrm{J}}(n)(t) + B(t) + L^{(1)}(n,t) + L^{(2)}(n,t) + L^{(3,1)}(n,t) + L^{(3,2)}(n,t), \end{aligned}$$

and for $t \in [0,T] \setminus G(n)$,

$$\begin{aligned} L(n,t) &= L\left(n, \left\lfloor \frac{t}{\Delta t(n)} \right\rfloor \Delta t(n)\right) + \gamma(n)\left(t - \left\lfloor \frac{t}{\Delta t(n)} \right\rfloor \Delta t(n)\right) \quad \text{and} \\ L^{\mathrm{D}}(n,t) &= L^{\mathrm{D}}\left(n, \left\lfloor \frac{t}{\Delta t(n)} \right\rfloor \Delta t(n)\right). \end{aligned}$$

Combining these expressions we discover that for $t \in G(n)$,

$$L^{\mathrm{R}}(n,t) = \left(\gamma^{\mathrm{J}}(n) - \gamma(n)\right)t + L^{(1)}(n,t) + L^{(2)}(n,t) + L^{(3,1)}(n,t),$$

and for $t \in [0,T] \setminus G(n)$,

$$L^{\mathrm{R}}(n,t) = L^{\mathrm{R}}\left(n, \left\lfloor \frac{t}{\Delta t(n)} \right\rfloor \Delta t(n)\right) + \gamma(n)\left(t - \left\lfloor \frac{t}{\Delta t(n)} \right\rfloor \Delta t(n)\right).$$

The triangle inequality then gives that

$$\begin{aligned}\sup_{t\in[0,T]} \|L^{\mathrm{R}}(n,t)\| \le \|\gamma^{\mathrm{J}}(n) - \gamma(n)\|\, T + \|\gamma(n)\|\Delta t(n) + \sup_{t\in[0,T]} \left\|L^{(1)}(n,t)\right\| \\ + \sup_{t\in[0,T]} \left\|L^{(2)}(n,t)\right\| + \sup_{t\in[0,T]} \left\|L^{(3,1)}(n,t)\right\|.\end{aligned}$$

The terms containing $\gamma(n)$ vanish as $n \to \infty$ by (3.3.4), Lemma 3.2.3, and (3.2.17). The remainder of the terms on the right hand side tend to zero in probability as $n \to \infty$ by (3.3.4) and Lemmas 3.2.1 and 3.2.2, proving this lemma. □

Proof of Theorem 3.3.1. We have that

$$\begin{aligned}&\rho\left(L^{\mathrm{D}}(n,\cdot), L(\cdot)\right) \\ &= \inf_{\lambda\in\Lambda}\left\{\sup_{t\in[0,T]} \|L^{\mathrm{D}}(n,t) - L(\lambda(t))\| + \sup_{t\in[0,T]} |\lambda(t) - t|\right\} \\ &\le \inf_{\lambda\in\Lambda}\left\{\sup_{t\in[0,T]} \|L^{\mathrm{R}}(n,t)\| + \sup_{t\in[0,T]} \|L(n,t) - L(\lambda(t))\| + \sup_{t\in[0,T]} |\lambda(t) - t|\right\} \\ &= \sup_{t\in[0,T]} \|L^{\mathrm{R}}(n,t)\| + \rho(L(n,\cdot), L(\cdot)) \xrightarrow{\mathrm{P}} 0, \quad n \to \infty,\end{aligned}$$

by (3.3.4), Lemma 3.3.1, and Theorem 3.2.1(i). This proves part (i).

$L^D(n,t)$ inherits independent increments from $L(t)$. Part (ii) therefore follows from part (i) and Lemma 3.2.7.

Let $\varepsilon, \delta > 0$ and $\sigma, \tau \in \mathcal{S}^{\mathrm{D}}_{0,T}(n)$ with $\tau \in [\sigma, \sigma+\delta]$. $\mathcal{S}^{\mathrm{D}}_{0,T}(n)$ is the set of $\mathbb{F}^{L^{\mathrm{D}}}(n)$-stopping times in $[0,T]$ and $\mathbb{F}^{L^{\mathrm{D}}}(n) = \{\mathcal{F}^{L^{\mathrm{D}}}_t(n)\}_{t\ge 0}$ is the filtration generated by

$L^{\mathrm{D}}(n,t)$. Now,

$$\begin{aligned}
&\mathrm{P}\left(\|L^{\mathrm{D}}(n,\tau) - L^{\mathrm{D}}(n,\sigma)\| \geq \varepsilon\right) \\
&\leq \mathrm{P}\left(\|L(n,\tau) - L(n,\sigma)\| \geq \frac{\varepsilon}{3}\right) + \mathrm{P}\left(\|L^{\mathrm{R}}(n,\tau)\| \geq \frac{\varepsilon}{3}\right) + \mathrm{P}\left(\|L^{\mathrm{R}}(n,\sigma)\| \geq \frac{\varepsilon}{3}\right).
\end{aligned}$$

The first term on the right hand side tends to zero as $n \to \infty$ and $\delta \to 0$ by (3.3.4) and the proof of Theorem 3.2.1(iii). Note that Theorem 3.2.1(iii) does not guarantee this result, but the proof of that theorem makes the more general assumption that the relevant stopping times belong to $\mathcal{S}^{\mathrm{D}}_{0,T}(n)$, so it can be applied here. For the remaining terms,

$$\mathrm{P}\left(\|L^{\mathrm{R}}(n,\tau)\| \geq \frac{\varepsilon}{3}\right) \leq \mathrm{P}\left(\sup_{t\in[0,T]} \|L^{\mathrm{R}}(n,t)\| \geq \frac{\varepsilon}{3}\right),$$

which tends to zero as $n \to \infty$ by Lemma 3.3.1. The same reasoning holds for σ. This proves part (iii). □

3.4 Convergence of American Option Prices

We now state our main results, starting with a convergence result for the price of an American option on a single asset under the skeleton approximation model.

Theorem 3.4.1. *Using the setup in Section 3.3, for $\pi(t)$ and $\pi^{D}(n,t)$ defined in* (2.1.2) *and* (3.3.3)*, with the additional assumptions that $d = 1$ and that the payoff function h is bounded and continuous in both its arguments,*

$$\lim_{n\to\infty} \pi^{D}(n,0) = \pi(0).$$

Proof. Apply Theorems 3.3.1 and 2.6.1 (Coquet and Toldo 2007, Theorem 5). □

While some authors (Këllezi and Webber 2004; Lord, Fang, Bervoets, and Oosterlee 2007, for example) have quite reasonably taken the price of an American option

to be the price of a Bermudan option in the limiting case of having an infinite number of exercise opportunities, Theorem 3.4.1 appears to be the first actual proof of this very intuitive result in a general Lévy model setting.

Remark 3.4.1. *The analogy to Theorem 3.4.1 for the first jump approximation model can be proved in an exactly analogous manner, but this proof is omitted as a similar result already exists in Szimayer and Maller (2007), Theorem 5.1.*

The following two theorems characterise the convergence of the American price processes under the skeleton and first jump approximation models to their continuous time counterpart. Theorem 3.4.2 is a generalisation of Theorem 4.3 in Maller, Solomon, and Szimayer (2006) to arbitrary dimensions, but differs as Theorem 4.3 in Maller et al. (2006) uses a binned first jump approximation model, as mentioned in Section 3.2, and makes certain Lipschitz continuity assumptions about the payoff function, rather than assuming a continuous bounded payoff function as we do here. Theorem 3.4.3 is a generalisation of Theorem 3.4.1 to arbitrary dimensions, but gives weak convergence in the Meyer-Zheng topology rather than convergence of $\pi^{\mathrm{D}}(n,0)$ to $\pi(0)$.

Theorem 3.4.2. *Under the assumptions of Theorem 3.2.1, for $\pi(t)$ and $\pi(n,t)$ defined in* (2.1.2) *and* (3.2.6)*, with the additional assumption that the payoff function h is continuous and bounded,*

$$\left(L(n,t), \mathrm{e}^{-rt}h(t,S(n,t)), \pi(n,t)\right) \overset{\text{MZ}}{\Longrightarrow} \left(L(t), \mathrm{e}^{-rt}h(t,S(t)), \pi(t)\right).$$

Theorem 3.4.3. *Using the setup in Section 3.3, for $\pi(t)$ and $\pi^{D}(n,t)$ defined in* (2.1.2) *and* (3.3.3)*, with the additional assumption that the payoff function h is continuous and bounded,*

$$\left(L^{D}(n,t), \mathrm{e}^{-rt}h\left(t,S^{D}(n,t)\right), \pi^{D}(n,t)\right) \overset{\text{MZ}}{\Longrightarrow} \left(L(t), \mathrm{e}^{-rt}h(t,S(t)), \pi(t)\right).$$

Proof of Theorems 3.4.2 and 3.4.3 will require the proof of two more theorems and a lemma, which are provided first.

The following two theorems, Theorems 3.4.4 and 3.4.5, are generalisations of Theorems 3.5 and 4.1 in Mulinacci and Pratelli (1998), respectively. Theorem 3.4.4 was inspired by Remark 3.9 in that paper. Theorems 3.5 and 4.1 from Mulinacci and Pratelli (1998) are reproduced in Section 2.6 for reference, and hypothesis (H), which is used in Theorem 3.4.4, was defined in Section 2.5. Note that although these generalisations are chiefly to accommodate a multivariate underlying asset, there is very little, in the theorems or their proofs, that is dependent on the dimensionality of the problem. Hence the proofs serve mainly to emphasise the ways in which the original theorems must be extended. Please refer to Mulinacci and Pratelli (1998) for a more detailed theoretical development.

Theorem 3.4.4. *Let $X(t)$ and $X(n,t)$, $n \in \mathbb{N}$, be random processes in $\mathbb{D}\left([0,T],\mathbb{R}^d\right)$ and $H(t)$, $J(t)$, $Y(t)$, $H(n,t)$, and $J(n,t)$ be random processes drawn from $\mathbb{D}([0,T],\mathbb{R})$ such that $J(n,t)$ and $J(t)$ are the Snell envelopes of $H(n,t)$ and $H(t)$ respectively. Let $\mathbb{F}^X$, $\mathbb{F}^{X,H}$, and $\mathbb{F}^{X,H}(n)$ be the natural filtrations of $X(t)$, $(X(t),H(t))$ and $(X(n,t),H(n,t))$ respectively. Define $\mathcal{S}_{0,T}^{X,H}(n)$ as the set of $\mathbb{F}^{X,H}(n)$-stopping times. Assume that*

(i) $\mathbb{F}^X = \mathbb{F}^{X,H}$,

(ii) $(X(n,t),H(n,t)) \overset{\text{S}}{\Longrightarrow} (X(t),H(t))$,

(iii) for all $n \in \mathbb{N}$ and $\tau(n) \in \mathcal{S}_{0,T}^{X,H}(n)$, $H(n,\tau(n))$ is uniformly integrable, and

(iv) $\{H(n,t)\}_{n\in\mathbb{N}}$ satisfies Aldous' criterion for tightness in expectation.

By considering if necessary a subsequence of $n \in \mathbb{N}$, we then have that for every $\tau(n) \in \mathcal{S}_{0,T}^{X,H}(n)$, $(X(n,t),H(n,t),J(n,t),\tau(n)) \overset{\text{MZ}}{\Longrightarrow} (X(t),H(t),Y(t),\theta)$ for some $(X(t),H(t),Y(t),\theta) \in \mathbb{D}\left([0,T],\mathbb{R}^d\right) \times [0,T]$. We let $\mathbb{F}^{X,H,Y,\theta}$ be the smallest right

continuous filtration to which $(X(t), H(t), Y(t))$ *is adapted and for which* θ *is an* $\mathbb{F}^{X,H,Y,\theta}$*-stopping time. Further assume that*

(v) for every $\tau(n) \in \mathcal{S}_{0,T}^{X,H}(n)$ *and all corresponding* $\mathbb{F}^{X,H,Y,\theta}$, $\left(X(t), \mathbb{F}^{X,H,Y,\theta}\right)$ *satisfies (H).*

Then we conclude that $(X(n,t), H(n,t), J(n,t)) \overset{\text{MZ}}{\Longrightarrow} (X(t), H(t), J(t))$.

For the next two proofs we adopt the notation of Mulinacci and Pratelli (1998) of denoting expectations on the probability spaces $\left(\Omega, \mathbb{F}^X, \mathrm{P}^X\right)$ and $\left(\Omega, \mathbb{F}^X(n), \mathrm{P}^X(n)\right)$, $n \in \mathbb{N}$, by E and E^n respectively, where P^X is the distribution of $X(t)$ and $\mathrm{P}^X(n)$ and $F^X(n)$ are the distributions and natural filtrations of $X(n,t)$.

Proof. Using assumption (i) we can say that $(X(n,t), H(n,t))$ is tight and, with paragraph (A.1) of Mulinacci and Pratelli (1998, p. 325), that $(X(n,t), H(n,t), J(n,t))$ is tight, for the Meyer-Zheng topology in both cases. By considering a subsequence if necessary, we can then suppose that $(X(n,t), H(n,t), J(n,t)) \overset{\text{MZ}}{\Longrightarrow} (X(t), H(t), Y(t))$. This allows us to say that $\mathrm{E}\, Y(t) \leq \liminf_{n\to\infty} \mathrm{E}^n J(n,t)$. Further, again by paragraph (A.1) in Mulinacci and Pratelli (1998), $Y(t)$ is a positive supermartingale, so

$$\mathrm{E}\, Y(t) = \sup_{c>0} \sup_{\delta\in(0,T-t)} \mathrm{E}\left[\frac{1}{\delta}\int_t^{t+\delta} (Y_s \wedge c)\,\mathrm{d}s\right], \tag{3.4.1}$$

and similarly for every $J(n,t)$. The functions $w(\cdot) \mapsto (1/\delta)\int_{\cdot}^{\cdot+\delta}(w(s)\wedge c)\,\mathrm{d}s$ are continuous on $\mathbb{D}([0,T],\mathbb{R})$ for the Meyer-Zheng topology, so, with (3.4.1), we have that $Y(\cdot) \mapsto \mathrm{E}\, Y(\cdot)$ is lower-semicontinuous for Meyer-Zheng convergence.

Now, by setting $\tau(n) = T$ in assumption (v), we obtain that $\left(X(t), \mathbb{F}^{X,H,Y}\right)$ satisfies hypothesis (H). Lemma 2.6 of Mulinacci and Pratelli (1998) and assumption (i) then give that the Snell envelope of $H(t)$ with respect to $\mathbb{F}^{X,H}$ coincides with that with respect to $\mathbb{F}^{X,H,Y}$, that is $J(t)$.

$Y(t)$ is an $\mathbb{F}^{X,H,Y}$-supermartingale with $Y(t) \geq X(t)$, $t \in [0,T]$, and so $Y(t) \geq J(t)$. We can now complete the proof by showing that $\mathrm{E}\, J(t) \geq \mathrm{E}\, Y(t)$. We proceed using proof by contradiction and assume that for some $t \in [0,T]$,

$$\varepsilon = \frac{\mathrm{E}\, Y(t) - \mathrm{E}\, J(t)}{4} > 0. \tag{3.4.2}$$

By (2.6.1), for all $n \in \mathbb{N}$, we can choose a stopping time $\tau(n) \in \mathcal{S}_{t,T}^{X,H}(n)$ such that $\mathrm{E}^n\, H(n,\tau(n)) \geq \mathrm{E}^n\, J(n,t) - \varepsilon$.

Lemmas 2.3 and 2.6 of Mulinacci and Pratelli (1998) give that the Snell envelopes of $H(t)$ on $\mathbb{D}\left([0,T],\mathbb{R}^{d+2}\right)$ with respect to $\mathbb{F}^{X,H,Y}$ and on $\mathbb{D}\left([0,T],\mathbb{R}^{d+2}\right)\times[0,T]$ with respect to $\mathbb{F}^{X,H,Y,\theta}$ have the same distribution. With the same tightness arguments as those employed in Mulinacci and Pratelli (1998, p. 316), we can then say that $(X(n,t),H(n,t),J(n,t),\tau(n)) \overset{\text{MZ}}{\Longrightarrow} (X(t),H(t),Y(t),\theta)$.

By Lemmas 3.3 and 3.4 from Mulinacci and Pratelli (1998), which rely on assumptions (iv) and (iii) respectively, we can find $\gamma, C > 0$ and $n_0 \in \mathbb{N}$ such that for $\delta \in (0,\gamma)$, $c > C$, and $n > n_0$,

$$\mathrm{E}^n\left[\frac{1}{\delta}\int_{\tau(n)}^{\tau(n)+\delta}(H(n,s)\wedge c)\,\mathrm{d}s\right] \geq \mathrm{E}^n\, H(n,\tau(n)) - 2\varepsilon \geq \mathrm{E}^n\, J(n,t) - 3\varepsilon.$$

Then, allowing $n \to \infty$,

$$\mathrm{E}\left[\frac{1}{\delta}\int_{\theta}^{\theta+\delta}(H(s)\wedge c)\,\mathrm{d}s\right] \geq \mathrm{E}\, Y(t) - 3\varepsilon.$$

Finally, letting $\delta \to 0$, we obtain that $\mathrm{E}(H(\theta)\wedge c) \geq \mathrm{E}\, Y(t) - 3\varepsilon$, so $\mathrm{E}\, J(t) \geq \mathrm{E}\, H(\theta) \geq \mathrm{E}\, Y(t) - 3\varepsilon$, which contradicts (3.4.2). Thus, $Y(t)$ coincides with $J(t)$. □

Theorem 3.4.5. *Define the random processes $X(t), X(n,t) \in \mathbb{D}\left([0,T],\mathbb{R}^d\right)$, $n \in \mathbb{N}$, and $H(n,t), H(t), J(n,t), J(t) \in \mathbb{D}([0,T],\mathbb{R})$ and the random variable $\theta \in [0,T]$. Let $\mathbb{F}^X(n) = \left(\mathcal{F}_t^X(n)\right)_{t\in[0,t]}$ be the natural filtration of $X(n,t)$; let $\mathbb{F}^{X,H,J}(n)$ be the natural filtration of $(X(n,t),H(n,t),J(n,t))$; let $\tau(n)$ be a $\mathbb{F}^X(n)$-stopping time; and*

let $\mathbb{F}^{X,H,J,\tau}(n)$ *be the smallest right continuous filtration to which* $(X(n,t), H(n,t), J(n,t))$ *is adapted and for which* $\tau(n)$ *is a* $\mathbb{F}^{X,H,J,\tau}(n)$*-stopping time. Assume that* $\mathbb{F}^X(n) = \mathbb{F}^{X,H,J}(n)$ *and* $(X(n,t), H(n,t), J(n,t), \tau(n)) \overset{\text{MZ}}{\Longrightarrow} (X(t), H(t), J(t), \theta)$*. Also define* $\mathbb{F}^{X,H,J,\theta} = \left(\mathcal{F}_t^{X,H,J,\theta}\right)_{t\in[0,T]}$ *as the smallest right continuous filtration to which* $(X(t), H(t), J(t))$ *is adapted and for which* θ *is an* $\mathbb{F}^{X,H,J,\theta}$*-stopping time. Suppose, for all* $n \in \mathbb{N}$ *and* $0 \leq s < t \leq T$*, that* $(X(n,t) - X(n,s))$ *is independent of* $\mathcal{F}_s^X(n)$*. Then* $(X(t) - X(s))$ *is independent of* $\mathcal{F}_s^{X,H,J,\theta}$*.*

Proof. From the stopping times $\tau(n)$ and θ define the random processes $Z(n,s;\omega) = \mathbf{1}_{\{s\geq\tau(n;\omega\}}$ and $Z(s;\omega) = \mathbf{1}_{\{s\geq\theta(\omega)\}}$. As shown in Mulinacci and Pratelli (1998), $Z(n,t) \overset{\text{MZ}}{\Longrightarrow} Z(t)$ if and only if $\tau(n)$ converges in distribution to θ, and $(X(n,t), H(n,t), J(n,t), Z(n,t)) \overset{\text{MZ}}{\Longrightarrow} (X(t), H(t), J(t), Z(t))$.

Consider a countable dense subset $\mathcal{T} \subset [0,T]$ and a subsequence such that the finite dimensional distributions of $(X(n,t), H(n,t), J(n,t), Z(n,t))_{t\in\mathcal{T}}$ converge to those of $(X(t), H(t), J(t), Z(t))_{t\in\mathcal{T}}$. Note that $\mathbb{F}^X(n) = \mathbb{F}^{X,H,J}(n) = \mathbb{F}^{X,H,J,\tau}(n)$ as $\tau(n)$ is an $\mathbb{F}^X(n)$-stopping time. For fixed $t_1 < \ldots < t_k < s < t \in \mathcal{T}$, a bounded continuous function $g : \mathbb{R}^{(d+3)k} \mapsto \mathbb{R}$, and all $n \in \mathbb{N}$,

$$
\begin{aligned}
&\mathrm{E}^n\left[\exp\left\{\mathrm{i}\sum_{j=1}^{d} u_j(X_j(n,t) - X_j(n,s))\right\}\right. \\
&\qquad \times g(X(n,t_1),\ldots,X(n,t_k),H(n,t_1),\ldots,H(n,t_k), \\
&\qquad\quad \left. J(n,t_1),\ldots,J(n,t_k),Z(n,t_1),\ldots,Z(n,t_k))\right] \\
&= \mathrm{E}^n\left[\exp\left\{\mathrm{i}\sum_{j=1}^{d} u_j(X_j(n,t) - X_j(n,s))\right\}\right] \\
&\qquad \times \mathrm{E}^n[g(X(n,t_1),\ldots,X(n,t_k),H(n,t_1),\ldots,H(n,t_k), \\
&\qquad\quad J(n,t_1),\ldots,J(n,t_k),Z(n,t_1),\ldots,Z(n,t_k))].
\end{aligned}
$$

This equality continues to hold as $n \to \infty$. Using the same monotone class argument

as that employed in Mulinacci and Pratelli (1998), we deduce that

$$\begin{aligned} &\mathrm{E}\left(\exp\left\{\mathrm{i}\sum_{j=1}^{d} u_j(X_j(n,t) - X_j(n,s))\right\}\middle|\mathcal{G}_s\right) \\ &= \mathrm{E}\left(\exp\left\{\mathrm{i}\sum_{j=1}^{d} u_j(X_j(n,t) - X_j(n,s))\right\}\right), \end{aligned}$$

where $\mathcal{G}_s$ is the σ-field generated by the random variables $(X(r), H(r), J(r), Z(r))$ with $r < s$. Thus $(X(t) - X(s))$ is independent of $\mathcal{G}_s$. As $\mathcal{F}_s^{X,H,J,\theta} = \mathcal{F}_s^{X,H,J,Z} = \bigcap_{r>s} \mathcal{G}_r$ and by the right continuity of paths, we have the required result. □

The following lemma, newly derived here, will be useful.

Lemma 3.4.1. *Let $X(t)$ and $\{X(n,t)\}_{n\in\mathbb{N}}$ be a process and sequence of processes in $\mathbb{D}\left([0,T],\mathbb{R}^d\right)$. Let f be a continuous function from $[0,T]\times\mathbb{R}^d$ to $\mathbb{R}^{d'}$. If $\{X(n,t)\}_{n\in\mathbb{N}}$ satisfies Aldous' criterion for tightness in probability and $\rho(X(n,\cdot), X(\cdot)) \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$, then*

(i) $\rho(f(\cdot, X(n,\cdot)), f(\cdot, X(\cdot))) \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$ and

(ii) $\{f(t, X(n,t))\}_{n\in\mathbb{N}}$ satisfies Aldous' criterion for tightness in probability.

Further, if $\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \|f(t,x)\| < \infty$,

(iii) $\{f(t, X(n,t))\}_{n\in\mathbb{N}}$ satisfies Aldous' criterion for tightness in expectation.

Proof. Define the event

$$A(n,c) = \left\{\sup_{t\in[0,T]} \|X(t)\| \le c \wedge \sup_{t\in[0,T]} \|X(n,t)\| \le c\right\},$$

for $n \in \mathbb{N}$ and $c > 0$. Also take $\lambda(n,t) \in \Lambda$. The Heine-Cantor theorem states that a continuous function on a compact metric space is uniformly continuous. So, on the set $A(n,c)$, for all $\varepsilon > 0$, there exists a $\eta > 0$ such that $\sup_{t\in[0,T]} \|(t, X(n,t)) -$

$(\lambda(n,t), X(\lambda(n,t)))\| < \eta$ implies that

$$\sup_{t\in[0,T]} \|f(t, X(n,t)) - f(\lambda(n,t), X(\lambda(n,t)))\| < \varepsilon.$$

Now,

$$\begin{aligned}
&\mathrm{P}\left(\sup_{t\in[0,T]} \|f(t, X(n,t)) - f(\lambda(n,t), X(\lambda(n,t)))\| \geq \varepsilon\right) \\
&\leq \mathrm{P}\left(\sup_{t\in[0,T]} \|f(t, X(n,t)) - f(\lambda(n,t), X(\lambda(n,t)))\| \geq \varepsilon \wedge A(n,c)\right) \\
&\quad + \mathrm{P}\left(\overline{A(n,c)}\right) \\
&\leq \mathrm{P}\left(\sup_{t\in[0,T]} \|(t, X(n,t)) - (\lambda(n,t), X(\lambda(n,t)))\| \geq \eta \wedge A(n,c)\right) \\
&\quad + \mathrm{P}\left(\sup_{t\in[0,T]} \|X(t)\| > c \vee \sup_{t\in[0,T]} \|X(n,t)\| > c\right) \\
&\leq \mathrm{P}\left(\sup_{t\in[0,T]} |t - \lambda(n,t)| \geq \frac{\eta}{2}\right) + \mathrm{P}\left(\sup_{t\in[0,T]} \|X(n,t) - X(\lambda(n,t))\| \geq \frac{\eta}{2}\right) \\
&\quad + \mathrm{P}\left(\sup_{t\in[0,T]} \|X(t)\| > c\right) + \mathrm{P}\left(\sup_{t\in[0,T]} \|X(n,t) - X(\lambda(n,t))\| > \frac{c}{2}\right) \\
&\quad + \mathrm{P}\left(\sup_{t\in[0,T]} \|X(\lambda(n,t))\| > \frac{c}{2}\right).
\end{aligned}$$

As $\rho(X(n,\cdot), X(\cdot)) \xrightarrow{\mathrm{P}} 0$, there exists a $\lambda(n,t) \in \Lambda$ such that

$$\begin{aligned}
&\mathrm{P}\left(\sup_{t\in[0,T]} |t - \lambda(n,t)| \geq \frac{\eta}{2}\right) + \mathrm{P}\left(\sup_{t\in[0,T]} \|X(n,t) - X(\lambda(n,t))\| \geq \frac{\eta}{2}\right) \\
&\quad + \mathrm{P}\left(\sup_{t\in[0,T]} \|X(n,t) - X(\lambda(n,t))\| > \frac{c}{2}\right)
\end{aligned}$$

tends to zero as $n \to \infty$. If we then allow $c \to \infty$,

$$\begin{aligned} &\mathrm{P}\left(\sup_{t\in[0,T]} \|X(t)\| > c\right) + \mathrm{P}\left(\sup_{t\in[0,T]} \|X(\lambda(n,t))\| > \frac{c}{2}\right) \\ &= \mathrm{P}\left(\sup_{t\in[0,T]} \|X(t)\| > c\right) + \mathrm{P}\left(\sup_{t\in[0,T]} \|X(t)\| > \frac{c}{2}\right) \end{aligned}$$

vanishes as $X(t)$ is càdlàg and therefore locally bounded. This proves part (i).

Turning now to part (ii), let $\mathcal{S}^{\mathrm{f}}_{0,T}(n)$ be the set of $\mathbb{F}^{\mathrm{f}}(n)$-stopping times and $\mathbb{F}^{\mathrm{f}}(n) = \{\mathcal{F}^{\mathrm{f}}_t(n)\}_{t\geq 0}$ be the filtration generated by $f(t, X(n,t))$. Define the set $B(n,c) = \left\{\sup_{t\in[0,T]} \|X(n,t)\| \leq c\right\}$ for $n \in \mathbb{N}$ and $c > 0$. For all paths in $B(n,c)$, the Heine-Cantor theorem again gives that for all $\varepsilon > 0$ there exists an $\eta > 0$ such that $\|(s, X(n,s)) - (t, X(n,t))\| < \eta$ implies that $\|f(s, X(n,s)) - f(t, X(n,t))\| < \varepsilon$ for all $s, t \in [0,T]$. So if we take $\delta > 0$ and $\sigma, \tau \in \mathcal{S}^{\mathrm{f}}_{0,T}(n)$ with $\tau \in [\sigma, \sigma + \delta]$,

$$\begin{aligned} &\mathrm{P}(\|f(\tau, X(n,\tau)) - f(\sigma, X(n,\sigma))\| \geq \varepsilon) \\ &\leq \mathrm{P}(\|f(\tau, X(n,\tau)) - f(\sigma, X(n,\sigma))\| \geq \varepsilon \wedge B(n,c)) + \mathrm{P}\left(\overline{B(n,c)}\right) \\ &\leq \mathrm{P}\left(\|X(n,\tau) - X(n,\sigma)\| \geq \frac{\eta}{2}\right) + \mathrm{P}\left(\tau - \sigma \geq \frac{\eta}{2}\right) \\ &\quad + \mathrm{P}\left(\sup_{t\in[0,T]} \|X(n,t) - X(\lambda(n,t))\| > \frac{c}{2}\right) \\ &\quad + \mathrm{P}\left(\sup_{t\in[0,T]} \|X(\lambda(n,t))\| > \frac{c}{2}\right), \end{aligned}$$

where $\lambda(n,t) \in \Lambda$. The last two terms in the final expression vanish as $n \to \infty$ followed by $c \to \infty$, as shown above. $\mathrm{P}(\tau - \sigma \geq \eta/2)$ becomes zero as soon as $\delta < \eta/2$. To deal with the remaining term, $f(t, X(n,t))$ is a function only of t and $X(n,\cdot)$ at t, so $\mathbb{F}^{\mathrm{f}}(n) \subset \mathbb{F}^{\mathrm{X}}(n)$, where $\mathbb{F}^{\mathrm{X}}(n)$ is the filtration generated by $X(n,t)$.

So,

$$\begin{aligned}&\sup_{\sigma,t\in\mathcal{S}^{\mathrm{f}}_{0,T}(n);\sigma\leq\tau\leq\sigma+\delta} \mathrm{P}\left(\|X(n,\tau)-X(n,\sigma)\|\geq\frac{\eta}{2}\right)\\ &\leq \sup_{\sigma,t\in\mathcal{S}^{X}_{0,T}(n);\sigma\leq\tau\leq\sigma+\delta} \mathrm{P}\left(\|X(n,\tau)-X(n,\sigma)\|\geq\frac{\eta}{2}\right).\end{aligned}$$

As $X(n,t)$ satisfies Aldous' criterion for tightness in probability, this expression tends to zero as $n\to\infty$ followed by $\delta\to 0$. Thus the requirements for part (ii) are met.

For part (iii), there exists a $C>0$ such that $\|f(t,x)\|<C$ for $(t,x)\in[0,T]\times\mathbb{R}^d$. Take $\varepsilon,\delta>0$ and $\sigma,\tau\in\mathcal{S}^{\mathrm{f}}_{0,T}(n)$ with $\tau\in[\sigma,\sigma+\delta]$, and write

$$\begin{aligned}&\mathrm{E}\,\|f(\tau,X(n,\tau))-f(\sigma,X(n,\sigma))\|\\ &\leq \mathrm{E}(\|f(\tau,X(n,\tau))-f(\sigma,X(n,\sigma))\|\,|\,\|f(\tau,X(n,\tau))-f(\sigma,X(n,\sigma))\|\geq\varepsilon)\\ &\quad+\mathrm{E}(\|f(\tau,X(n,\tau))-f(\sigma,X(n,\sigma))\|\,|\,\|f(\tau,X(n,\tau))-f(\sigma,X(n,\sigma))\|<\varepsilon)\\ &\leq 2C\,\mathrm{P}(\|f(\tau,X(n,\tau))-f(\sigma,X(n,\sigma))\|\geq\varepsilon)+\varepsilon.\end{aligned}$$

Therefore,

$$\begin{aligned}&\sup_{\sigma,t\in\mathcal{S}^{\mathrm{f}}_{0,T}(n);\sigma\leq\tau\leq\sigma+\delta} \mathrm{E}\,\|f(\tau,X(n,\tau))-f(\sigma,X(n,\sigma))\|\\ &\leq 2C\sup_{\sigma,t\in\mathcal{S}^{\mathrm{f}}_{0,T}(n);\sigma\leq\tau\leq\sigma+\delta} \mathrm{P}\left(\|X(n,\tau)-X(n,\sigma)\|\geq\varepsilon\right)+\varepsilon.\end{aligned}$$

The first term in the above expression was shown to disappear as $n\to\infty$ followed by $\delta\to 0$ in part (ii). ε can then be made arbitrarily small, proving part (iii). □

As mentioned above, Theorem 4.3 of Maller et al. (2006) bears some similarities to Theorem 3.4.2 in the one dimensional case. While the proofs of these theorems both rely on results from Mulinacci and Pratelli (1998) (or their generalisations), the way in which they apply those results differs.

Proof of Theorem 3.4.2. In the notation of Theorem 3.4.4, set $X(t) = L(t)$, $X(n,t) = L(n,t)$, $H(n,t) = \mathrm{e}^{-rt}h(t,S(n,t))$, $H(t) = \mathrm{e}^{-rt}h(t,S(t))$, $J(n,t) = \pi(n,t)$, and $J(t) = \pi(t)$. Assumption (i) of Theorem 3.4.4 holds as $H(t)$ is a function only of t and $L(t)$. Noting that $H(n,t)$ and $H(t)$ are continuous, bounded functions of $L(n,t)$ and $L(t)$ respectively, assumptions (ii) and (iv) follow from Theorem 3.2.1 in conjunction with Lemma 3.4.1. Assumption (iii) holds as $H(n,t)$ is bounded.

For assumption (v), $\left(X(t), \mathbb{F}^{X,H,J,\theta}\right)$ satisfies hypothesis (H) if $X(t)$ is Markovian with respect to each $\mathbb{F}^{X,H,J,\theta}$. This requirement is given by Theorem 3.4.5 if $(X(n,t) - X(n,s))$ is independent of $\mathcal{F}_s^X(n)$ for all $n \in \mathbb{N}$ and $0 \leq s < t \leq T$ and if $\mathbb{F}^{X,H,J}(n) = \mathbb{F}^X(n)$. These two conditions are true as $X(n,t) = L(n,t)$ is a sum of independent variables and $H(n,t)$ and $J(n,t)$ can be formulated as functions only of t and $X(n,t)$.

Theorem 3.4.4 now gives the result. □

The proof below follows the same lines as that above.

Proof of Theorem 3.4.3. In the notation of Theorem 3.4.4, set $X(t) = L(t)$, $X(n,t) = L^{\mathrm{D}}(n,t)$, $H(n,t) = \mathrm{e}^{-rt}h\left(t, S^{\mathrm{D}}(n,t)\right)$, $H(t) = \mathrm{e}^{-rt}h(t,S(t))$, $J(n,t) = \pi^{\mathrm{D}}(n,t)$, and $J(t) = \pi(t)$. Assumption (i) of Theorem 3.4.4 holds as $H(t)$ is a function only of t and $L^{\mathrm{D}}(t)$. Noting that $H(n,t)$ and $H(t)$ are continuous, bounded functions of $L(n,t)$ and $L(t)$ respectively, assumptions (ii) and (iv) follow from Theorem 3.3.1 in conjunction with Lemma 3.4.1. Assumption (iii) holds as $H(n,t)$ is bounded.

For assumption (v), $\left(X(t), \mathbb{F}^{X,H,J,\theta}\right)$ satisfies hypothesis (H) if $X(t)$ is Markovian with respect to each $\mathbb{F}^{X,H,J,\theta}$. This requirement is given by Theorem 3.4.5 if $(X(n,t) - X(n,s))$ is independent of $\mathcal{F}_s^X(n)$ for all $n \in \mathbb{N}$ and $0 \leq s < t \leq T$ and if $\mathbb{F}^{X,H,J}(n) = \mathbb{F}^X(n)$. These two conditions are true as $X(n,t) = L^{\mathrm{D}}(n,t)$ inherits independent increments from the Lévy process from which it is derived and $H(n,t)$ and $J(n,t)$ can be formulated as functions only of t and $X(n,t)$.

Theorem 3.4.4 now gives the result. □

3.5 An Example Application of Theorem 3.3.1

The following theorem demonstrates how Theorem 3.3.1 can be used to answer questions about whether prices gotten using existing lattice models converge asymptotically to the correct price under the continuous time model. The lattice model chosen in this example is the multinomial tree model based on transition probabilities that is given in Këllezi and Webber (2004). We show how this multinomial model does converge after slight modifications and under certain conditions. See Remark 2.3.1 for details of the modifications.

Assume the setup given in Section 3.3. Define the additional quantities $\Delta l(n) \in \mathbb{R}^+$, $u(n), d(n) \in \mathbb{N}$, and $\mathcal{K}(n) = \{-d(n), \ldots, u(n)\}$. Also take the random processes $\widetilde{L}(n,t) \in \mathbb{D}([0,T], \mathbb{R})$, $\widetilde{S}(n,t) = S(0)\mathrm{e}^{\widetilde{L}(n,t)}$, and

$$\widetilde{\pi}(n,t) = \operatorname*{ess\,sup}_{\tau \in \mathcal{S}_{t,T}^{\widetilde{L}}} \mathrm{E}\left(\mathrm{e}^{-r(\tau-t)} h(\tau, \widetilde{S}(n,\tau)) \middle| \mathcal{F}_t^{\widetilde{L}}(n) \right), \quad t \in [0,T],$$

where $\mathbb{F}^{\widetilde{L}}(n) = \left(\mathcal{F}_t^{\widetilde{L}}(n)\right)_{t\in[0,T]}$ is the natural filtration of $\widetilde{L}(n,t)$ and $\mathcal{S}_{t,T}^{\widetilde{L}}$ is the set of $\mathbb{F}^{\widetilde{L}}(n)$-stopping times in $[t,T]$.

Further constrain $\widetilde{L}(n,t)$ in terms of the skeleton approximation $L^{\mathrm{D}}(n,t)$:

$$\Delta\widetilde{L}(n,t) = \Delta l(n) \operatorname*{arg\,min}_{k \in \mathcal{K}(n)} \left| k\Delta l(n) - L^{\mathrm{D}}(n,t) \right|, \quad t \in [0,T],$$

and set $\widetilde{L}(n,0) = 0$.

Theorem 3.5.1. *Assume the above definitions. Further assume that the payoff function $h : [0,T] \times \mathbb{R}^+ \mapsto \mathbb{R}^+$ is bounded and continuous. If, as $n \to \infty$,*

$$\frac{\Delta l(n)}{\Delta t(n)} \to 0$$

and

$$\frac{1}{\Delta t(n)} \mathrm{P}\left(\Delta L^{D}(n, \Delta t(n)) \notin (-d(n)\Delta l(n), u(n)\Delta l(n))\right) \to 0,$$

then

$$\lim_{n\to\infty} \widetilde{\pi}(n, 0) = \pi(0).$$

We prove this theorem using the following two lemmas.

Lemma 3.5.1. *Under the assumptions of Theorem 3.5.1,*

$$\sup_{t\in[0,T]} \left|\widetilde{L}(n,t) - L^{D}(n,t)\right| \xrightarrow{\mathrm{P}} 0, \quad n \to 0.$$

Proof. Define the event $A = \{\Delta L^{\mathrm{D}}(n,t) \in (-d(n)\Delta l(n), u(n)\Delta l(n)), t \in [0,T]\}$ and its complement $\overline{A}$. For any $\varepsilon > 0$,

$$\mathrm{P}\left(\sup_{t\in[0,T]} \left|\widetilde{L}(n,t) - L^{\mathrm{D}}(n,t)\right| \geq \varepsilon\right) \leq \mathrm{P}\left(\sup_{t\in[0,T]} \left|\widetilde{L}(n,t) - L^{\mathrm{D}}(n,t)\right| \geq \varepsilon \middle| A\right) + \mathrm{P}(\overline{A}).$$

The first term in the final expression above tends to zero as $n \to \infty$ because, if we assume that A is true,

$$\sup_{t\in[0,T]} \left|\widetilde{L}(n,t) - L^{\mathrm{D}}(n,t)\right| \leq N(n)\Delta l(n) = T\frac{\Delta l(n)}{\Delta t(n)} \to 0.$$

For the second term, as $L^{\mathrm{D}}(n,t)$ has independent increments,

$$\begin{aligned}\mathrm{P}(\overline{A}) &\leq N(n)\,\mathrm{P}\left(\Delta L^{\mathrm{D}}(n, \Delta t(n)) \notin (-d(n)\Delta l(n), u(n)\Delta l(n))\right) \\ &= \frac{T}{\Delta t(n)} \mathrm{P}\left(\Delta L^{\mathrm{D}}(n, \Delta t(n)) \notin (-d(n)\Delta l(n), u(n)\Delta l(n))\right) \to 0\end{aligned}$$

as $n \to \infty$, which concludes the proof. □

Lemma 3.5.2. *Under the assumptions of Theorem 3.5.1, As $n \to \infty$,*

(i) $\rho\left(\widetilde{L}(n,\cdot), L(\cdot)\right) \xrightarrow{\mathrm{P}} 0$,

(ii) $\mathbb{F}^{\widetilde{L}}(n) \xrightarrow{\mathrm{W}} \mathbb{F}^{L}$, *and*

(iii) $\left\{\widetilde{L}(n,t)\right\}_{n\in\mathbb{N}}$ *fulfills Aldous' criterion for tightness in probability.*

Proof. Let Λ be the set of strictly increasing continuous functions $\lambda : [0,T] \mapsto [0,T]$ with $\lambda(0) = 0$ and $\lambda(T) = T$. We have

$$\begin{aligned}\rho\left(\widetilde{L}(n,\cdot), L(\cdot)\right) &= \inf_{\lambda\in\Lambda}\left\{\sup_{t\in[0,T]}\left|\widetilde{L}(n,t) - L(\lambda(t))\right| + \sup_{t\in[0,T]}|\lambda(t)-t|\right\}\\ &\le \inf_{\lambda\in\Lambda}\left\{\sup_{t\in[0,T]}|L^{\mathrm{D}}(n,t) - L(\lambda(t))| + \sup_{t\in[0,T]}|\lambda(t)-t|\right\}\\ &\quad + \sup_{t\in[0,T]}\left|\widetilde{L}(n,t) - L^{\mathrm{D}}(n,t)\right|.\end{aligned}$$

The first term on the right-hand side is just $\rho((L^{\mathrm{D}}(n,\cdot), L(\cdot))$, which tends to zero in probability by Theorem 3.3.1(i). The second term also tends to zero in probability by Lemma 3.5.1, proving part (i).

As $\widetilde{L}(n,t)$ possesses independent increments, by part (i) and Theorem 3.2.7, part (ii) follows.

Remember that $\mathbb{F}^{\widetilde{L}}(n)$ is the natural filtration of $\widetilde{L}(n,t)$ and $\mathcal{S}^{\widetilde{L}}_{0,T}(n)$ is the set of $\mathbb{F}^{\widetilde{L}}(n)$-stopping times. Take $\mathbb{F}^{L^{\mathrm{D}}}(n)$ to be the natural filtration of $L^{\mathrm{D}}(n,t)$ and $\mathcal{S}^{L^{\mathrm{D}}}_{0,T}(n)$ to be the set of $\mathbb{F}^{L^{\mathrm{D}}}(n)$-stopping times. Now,

$$\begin{aligned}&\lim_{\delta\searrow 0}\limsup_{n\to\infty}\sup_{\sigma,\tau\in\mathcal{S}^{\widetilde{L}}_{0,T}(n);\sigma\le\tau\le\sigma+\delta}\mathrm{P}\left(\left|\widetilde{L}(n,\tau) - \widetilde{L}(n,\sigma)\right| \ge \varepsilon\right)\\ &\le \lim_{\delta\searrow 0}\limsup_{n\to\infty}\sup_{\sigma,\tau\in\mathcal{S}^{\widetilde{L}}_{0,T}(n);\sigma\le\tau\le\sigma+\delta}\left[\mathrm{P}\left(\left|\widetilde{L}(n,\tau) - \widetilde{L}(n,\sigma)\right| \ge \frac{\varepsilon}{3}\right)\right.\\ &\quad \left. + \mathrm{P}\left(\left|\widetilde{L}(n,\tau) - L^{\mathrm{D}}(n,\tau)\right| \ge \frac{\varepsilon}{3}\right) + \mathrm{P}\left(\left|\widetilde{L}(n,\sigma) - L^{\mathrm{D}}(n,\sigma)\right| \ge \frac{\varepsilon}{3}\right)\right]\end{aligned}$$

$$\leq \lim_{\delta \searrow 0} \limsup_{n \to \infty} \sup_{\sigma,\tau \in \mathcal{S}_{0,T}^{L^{\mathrm{D}}}(n);\sigma \leq \tau \leq \sigma+\delta} \mathrm{P}\left(\left|\widetilde{L}(n,\tau) - \widetilde{L}(n,\sigma)\right| \geq \frac{\varepsilon}{3}\right)$$
$$+ 2 \limsup_{n \to \infty} \mathrm{P}\left(\sup_{t \in [0,T]} \left|\widetilde{L}(n,t) - L^{\mathrm{D}}(n,t)\right| \geq \frac{\varepsilon}{3}\right) = 0,$$

where we have used that, by construction, $\mathcal{S}_{0,T}^{\widetilde{L}}(n) \subset \mathcal{S}_{0,T}^{L^{\mathrm{D}}}(n)$, Theorem 3.3.1(iii), and Lemma 3.5.1. We therefore have part (iii). □

Proof of Theorem 3.5.1. Take $\mathcal{S}_{0,T}(n)$ to be the set of stopping times with respect to the natural filtration of $L(t)$. Note that

$$\widetilde{\pi}(n,0) = \sup_{\tau \in \mathcal{S}_{0,T}^{\widetilde{L}}(n)} \mathrm{E}\left(h\left(\tau, S(0)\mathrm{e}^{\widetilde{L}(n,\tau)}\right)\right) = \sup_{\tau \in \mathcal{S}_{0,T}^{\widetilde{L}}(n)} \mathrm{E}\left(f\left(\tau, \widetilde{L}(n,\tau)\right)\right)$$
$$\text{and} \quad \pi(0) = \sup_{\tau \in \mathcal{S}_{0,T}} \mathrm{E}\left(h\left(\tau, S(0)\mathrm{e}^{L(\tau)}\right)\right) = \sup_{\tau \in \mathcal{S}_{0,T}} \mathrm{E}\left(f\left(\tau, L(\tau)\right)\right)$$

for some bounded and continuous function $f : [0,T] \times \mathbb{R}^+ \mapsto \mathbb{R}^+$. The required result is now given by Theorem 2.6.1, under its conditions, which are shown to be satisfied by Lemma 3.5.2. □

3.6 Directions for Future Work

The convergence of an American option price obtained using a multinomial model to its continuous time limit was first proved in Maller et al. (2006). Convergence in this instance was weak convergence in the Meyer-Zheng topology. This work was generalised for an arbitrary number of underlying assets in Theorems 3.4.2 and 3.4.3, and required the generalisation of several results from Mulinacci and Pratelli (1998). In Szimayer and Maller (2007), two of the authors of Maller et al. (2006) presented another convergence result for the multinomial model described in Maller et al. (2006), using a new theoretical result made available in Coquet and Toldo (2007). In conjunction with Theorem 3.3.1, the same result from Coquet and Toldo (2007) was used in Theorem 3.4.1 to prove convergence for prices obtained with the

skeleton approximation model for American options on a single asset.

A possible direction for future enquiry would be to generalise Theorem 3.4.1 for rainbow options, or similarly to generalise the convergence results of Szimayer and Maller (2007) for arbitrary dimensions. As some results in Mulinacci and Pratelli (1998) were generalised to prove Theorems 3.4.2 and 3.4.3, this would require a multidimensional generalisation of the results in Coquet and Toldo (2007). More specifically, the assumption of a single underlying asset in Theorem 3.4.1 could be lifted if the result given in Theorem 2.6.1, which is Theorem 5 in Coquet and Toldo (2007), were valid for multidimensional stochastic processes. The analogous result for the first jump approximation could be similarly obtained. Generalisation of Theorem 2.6.1 is beyond the scope of this thesis and is left as a topic for further investigation.

Remark 3.6.1. *Theorem 3.4.1 could be proved for American rainbows using Theorem 2.6.1 in its current form if, in the notation of Section 3.3,*

(i) $\rho(h(\cdot, S^D(n,\cdot)), h(\cdot, S^D(\cdot))) \xrightarrow{\mathrm{P}} 0$,

(ii) $\{h(t, S^D(n,t))\}_{n\in\mathbb{N}}$ *satisfied Aldous' criterion for tightness in probability, and*

(iii) $\mathbb{F}^h(n) \xrightarrow{\mathrm{W}} \mathbb{F}^h$ *as* $n \to \infty$ *or* $\mathbb{F}^h(n) \subset \mathbb{F}^h$ *for all* $n \in \mathbb{N}$,

where $\mathbb{F}^h(n) = \left(\mathcal{F}_t^h(n)\right)_{t\in[0,T]}$, $n \in \mathbb{N}$, *and* $\mathbb{F}^h = (\mathcal{F}_t^h)_{t\in[0,T]}$ *are the filtrations generated by* $h(t, S^D(n,t))$ *and* $h(t, S(t))$ *respectively. Conditions (i) and (ii) are given by Theorem 3.3.1(i) and (iii) in conjunction with Lemma 3.4.1. However, condition (iii) does not follow from weak convergence or nestedness between the filtrations generated by* $S^D(n,t)$ *and* $S(t)$ *or, equivalently,* $L^D(n,t)$ *and* $L(t)$*, which we now demonstrate.*

Take a process $X(t)$ *and a sequence of processes* $\{X(n,t)\}_{n\in\mathbb{N}}$*, all drawn from* $\mathbb{D}\left([0,T], \mathbb{R}^d\right)$*. Let the filtrations* $(\mathcal{F}_t^X)_{t\in[0,T]} = \mathbb{F}^X$ *and* $\left(\mathcal{F}_t^X(n)\right)_{t\in[0,T]} = \mathbb{F}^X(n)$ *for* $n \in \mathbb{N}$ *be the filtrations generated by* $X(t)$ *and* $\{X(n,t)\}_{n\in\mathbb{N}}$*. Also let* $f : [0,T]\times\mathbb{R}^d \mapsto$

$\mathbb{R}$ *be a continuous and bounded function and* $(\mathcal{F}_t^f)_{t\in[0,T]} = \mathbb{F}^f$ *and* $\left(\mathcal{F}_t^f(n)\right)_{t\in[0,T]} = \mathbb{F}^f(n)$ *for* $n \in \mathbb{N}$ *be the filtrations generated by* $f(t, X(t))$ *and* $f(t, X(n,t))$. *Now,*

(i) $\mathbb{F}^X(n) \xrightarrow{\text{w}} \mathbb{F}^X$ *does not imply that* $\mathbb{F}^f(n) \xrightarrow{\text{w}} \mathbb{F}^f$ *and*

(ii) $\mathbb{F}^X(n) \subset \mathbb{F}^X$ *does not imply that* $\mathbb{F}^f(n) \subset \mathbb{F}^f$.

The counterexample that demonstrates (i) is constructed as follows. Take a process $Y(t) \in \mathbb{D}([0,T], \mathbb{R})$ *and define*

$$X(n,t) = \left(Y(t), \frac{Y(t)}{n}\right), \quad X(t) = (Y(t), 0), \quad \text{and} \quad f(t,(x_1,x_2)) = x_2.$$

Now $\mathbb{F}^X = \mathbb{F}^X(n) = \mathbb{F}^f(n)$, *but* $\mathbb{F}^f$ *is the trivial filtration. That is the application of* f *removes information from* $X(t)$ *without removing it from* $X(n,t)$.

The counterexample that supports (ii) proceeds thus. Choose an $n \in \mathbb{N}$. *Define two independent Bernoulli random variables* B_1 *and* B_2 *and set*

$$\begin{aligned} X(t) = B_1 \quad \text{and} \quad X(n,t) = 0 \qquad & \text{for} \quad t \in [0, T/2), \\ X(t) = B_1 + B_2 \quad \text{and} \quad X(n,t) = B_2 \qquad & \text{for} \quad t \in [T/2, T], \end{aligned}$$

$$\text{and} \quad f(t,x) = \begin{cases} 0, & x \in (-\infty, 0), \\ x, & x \in [0,1), \\ 1, & x \in [x, \infty). \end{cases}$$

We can now characterise the sample set $\Omega = \{\{0,1\} \times \{0,1\}\}$, *representing the possible outcomes of* B_1 *and* B_2. *Writing out the* σ*-algebras* $\mathcal{F}_t^X$, $\mathcal{F}_t^f$, $\mathcal{F}_t^X(n)$, *and* $\mathcal{F}_t^f(n)$, *for* $t \in \{0, T/2\}$ *reveals that* $\mathbb{F}^X(n) \subset \mathbb{F}^X$, *but* $\mathcal{F}_{\frac{T}{2}}^f(n) \nsubseteq \mathcal{F}_{\frac{T}{2}}^f$. *Again,* f *destroys information about* $X(t)$ *which survives in the process* $f(t, X(n,t))$.

There has not yet been any discussion as to why the first jump approximation model given in Section 3.2 does not bin jumps in the same way that the multinomial model of Maller et al. (2006) does, or of how one might implement a program to price

options using the skeleton approximation model. These issues, which are primarily regarding the efficiency and accuracy of numerical computation of prices, are addressed in the next chapter, where a simple example of the skeleton approximation model is implemented.

Chapter 4

Implementation

4.1 Algorithm

This chapter provides details of how one might use the results gotten in Chapter 3 to implement procedures for pricing American options. The price convergence results in Chapter 3 were deliberately general and can be used as the basis for many lattice pricing models. Indeed, they can be used to show convergence for several models already demonstrated in the literature, see Remarks 2.3.1, 2.4.1, and 2.4.2. The algorithm given here is kept as general as possible, and is applicable to both the first jump and skeleton approximation models, but to illustrate the concepts a simple example is explored: pricing an American put in one dimension using a VG process and the skeleton approximation model.

Theoretically, the Bellman equation, introduced in Chapter 2 as (2.4.2), solves the dynamic programming problem of pricing an American option under either of our approximation schemes. In the case of the first jump approximation model, so assuming the setup in Section 3.2, this equation becomes

$$\begin{aligned} &v(n,(i-1)\Delta t(n),x) \\ &= \max\left(h((i-1)\Delta t(n),x), \mathrm{e}^{-r\Delta t(n)}\,\mathrm{E}\left(v\left(n,i\Delta t(n),x\mathrm{e}^{L(n,\Delta t)}\right)\right)\right) \end{aligned} \tag{4.1.1}$$

for $i \in \{0, \ldots, N(n)-1\}$ with $v(n,T,x) = h(T,x)$. Here $v(n,\cdot,\cdot) : \{i\Delta t(n)\}_{i\in\mathbb{N}} \times \mathbb{R}^{+d} \mapsto \mathbb{R}^+$ is the approximate value function satisfying

$$v(n,t,x) = \sup_{\tau \in \mathcal{S}_{0,T-t}(n)} \mathrm{E}\left(\mathrm{e}^{-r\tau} h\left(t+\tau, x_1 \mathrm{e}^{L_1(n,\tau)}, \ldots, x_d \mathrm{e}^{L_d(n,\tau)}\right)\right),$$

so $v(n, i\Delta t(n), S(n, i\Delta t(n))) = \pi(n, i\Delta t(n))$. The challenge is to find an efficient means of calculating $v(n,0,S(0))$.

The first jump approximation model is a generalisation of the multinomial model of Maller, Solomon, and Szimayer (2006) and Szimayer and Maller (2007), which is described in Section 2.3. Whereas the multinomial model is prescriptive about the tree method to be used in evaluating (4.1.1), both the first jump and skeleton approximation models require that further decisions be made about how they might be implemented. We use the IL algorithm of Kargin (2005), defined in Algorithm 2.4.2, as a framework for these decisions. This algorithm calls for an approximation to the value function $v(n, i\Delta t(n), \cdot)$ to be constructed for each $i \in \{1, \ldots, N(n)\}$, which we denote $\widehat{v}(n,i,\cdot)$. As a step towards arriving at this approximation we define the approximate continuation value function

$$\widehat{c}(n,i,x) = \mathrm{e}^{-r\Delta t(n)} \mathrm{E}\left(\widehat{v}\left(n, i+1, x\mathrm{e}^{L(n,\Delta t)}\right)\right) + e_c(n,i,x), \tag{4.1.2}$$

for $x \in \mathbb{R}^{+d}$, where e_c is a function capturing the error in this numerical calculation. We use these functions to define another error function e_v:

$$\widehat{v}(n,i,x) = \max(h(i\Delta t(n), x), \widehat{c}(n,i,x)) + e_v(n,i,x), \quad x \in \mathbb{R}^{+d}. \tag{4.1.3}$$

We also define a finite grid $\mathcal{G}(n,i) \subset \mathbb{R}^+$, $i \in \{1, \ldots, N(n)\}$, which is not necessarily regular. Using this notation, Algorithm 2.4.2 can be rewritten as Algorithm 4.1.1.

The original multinomial model could be seen as a specific implementation of the first jump approximation model in one dimension using Algorithm 4.1.1. In this

Algorithm 4.1.1 The IL algorithm for pricing Americans, revised.

Generate $\mathcal{G}(n,i)$, $i \in \{1,\dots,N(n)\}$

Set $\widehat{v}(n,T,x) = h(T,x)$.

for $i = N(n)-1$ to 1 **do**

Fit $\widehat{v}(n,i,\cdot)$ to $\{(x, \max(h(i\Delta t(n),x), \widehat{c}(n,i,x))) : x \in \mathcal{G}(n,i)\}$

end for

return $\max(h(0,S(0)), \widehat{c}(n,0,S(0)))$

case, we would set

$$\mathcal{G}(n) = \{S(0)\mathrm{e}^{i\gamma(n)+k\Delta(n))}\}_{k\in\{-iM(n),\dots,iM(n)\}},$$

where $M(n) \in \mathbb{N}$, $\Delta(n) \in \mathbb{R}^+$, and we assume the setup in Section 3.2. We set the jump selection region to be

$$\mathcal{J}(n) = \left[-\left(M(n)+\frac{1}{2}\right)\Delta(n), -\frac{\Delta(n)}{2}\right) \cup \left(\frac{\Delta(n)}{2}, \left(M(n)+\frac{1}{2}\right)\Delta(n)\right].$$

In this context, $\widehat{v}(n,\cdot,\cdot)$ is only ever evaluated where it is calculated, so no approximation is required and $e_v(n,\cdot,\cdot) = 0$. If we make the simplifying assumption that $L(t)$ is a pure jump process, then

$$\widehat{c}(n,i,x) = \mathrm{e}^{-r\Delta t(n)} \sum_{k=-M(n)}^{M(n)} \widehat{v}\left(n, i+1, x\mathrm{e}^{\gamma(n)+k\Delta(n)}\right) p_k(n), \tag{4.1.4}$$

where

$$p_k(n) = \begin{cases} \int_{\left(k-\frac{1}{2}\right)\Delta(n)}^{\left(k+\frac{1}{2}\right)\Delta(n)} f_n^{\Pi}(y)\,\mathrm{d}y, & k \in \{-M(n),\dots,M(n)\}\setminus\{0\}, \\ \mathrm{e}^{-\Delta t(n)\Pi\{\mathcal{J}(n)\}}, & k = 0, \end{cases}$$

and $f_n^{\Pi}(y)\,\mathrm{d}y = \dfrac{1-\mathrm{e}^{-\Delta t(n)\Pi\{\mathcal{J}(n)\}}}{\Pi\{\mathcal{J}(n)\}}\Pi\{\mathrm{d}y\}$.

We can then write

$$
\begin{aligned}
|e_c(n,i-1,x)| = \mathrm{e}^{-r\Delta t(n)} \Bigg| & \sum_{k=-M(n)}^{M(n)} \widehat{v}\left(n,i+1,x\mathrm{e}^{\gamma(n)+k\Delta(n)}\right) p_k(n), \\
& - \int_{\mathcal{J}(n)} \widehat{v}\left(n,i+1,x\mathrm{e}^{\gamma(n)+y}\right) f_n^{\Pi}(y)\,\mathrm{d}y - \widehat{v}\left(n,i+1,x\mathrm{e}^{\gamma(n)}\right) p_0(n) \Bigg| \\
\leq \Delta^3(n) M(n) & \frac{\mathrm{e}^{-r\Delta t(n)}}{12} \max_{y\in\mathcal{J}(n)} \left| \frac{\partial^2}{\partial y^2} \widehat{v}\left(n,i+1,x\mathrm{e}^{\gamma(n)+y}\right) f_n^{\Pi}(y) \right. \\
& \left. + 2\frac{\partial}{\partial y} \widehat{v}\left(n,i+1,x\mathrm{e}^{\gamma(n)+y}\right) \frac{\mathrm{d}}{\mathrm{d}y} f_n^{\Pi}(y) \right|,
\end{aligned}
$$

assuming that $\widehat{v}(n,i,\cdot)$ and $f_n^{\Pi}(\cdot)$ are sufficiently smooth in the region of interest. While this assumption may not be realistic everywhere, see the discussion of smooth pasting in Section 2.3, the leading term of $\Delta^3(n)M(n)$ demonstrates that this numerical integration scheme could be made to converge faster. This rate of convergence is the same as that for the extended midpoint rule (Press, Teukolsky, Vetterling, and Flannery 1992, p. 135), which is a simple numerical quadrature rule that is very similar in form to (4.1.4). With reweighting of the sum in (4.1.4),

$$
\widehat{c}(n,i,x) = \mathrm{e}^{-r\Delta t(n)} \sum_{k=-M(n)}^{M(n)} w_k \widehat{v}\left(n,i+1,x\mathrm{e}^{\gamma(n)+k\Delta(n)}\right) p_k(n), \quad w_k \in \mathbb{R}^+,
$$

better rates of convergence could be achieved, say by applying a Newton-Cotes formula. See Press et al. (1992, Chapter 4) for a review of techniques in numerical integration.

The observation that the convergence of lattice methods can be accelerated by appropriately weighting the transition probabilities has been made before. The approach of Lord, Fang, Bervoets, and Oosterlee (2007), which was discussed in Section 2.4, uses a Newton-Cotes formula in this fashion. In that paper this idea was attributed to Andricopoulos, Widdicks, Duck, and Newton (2003), in which it is used to speed recombining tree methods in the Black-Scholes setting.

Here we have made this point to justify the use of the IL algorithm. The per-

formance of a regular lattice model can be improved by reweighting the transition probability mass function. Being able to choose where to evaluate $\widehat{v}(n,i,\cdot)$ when calculating $\widehat{c}(n,i-1,x)$ has the potential to bring further improvements in accuracy through the use of integration techniques such as Gaussian quadrature, adaptive integration, or variable transformation methods (Press et al. 1992, p. 147). The benefit of this approach is, however, limited by the error in the approximation $\widehat{v}(n,i,\cdot)$ and the latency introduced by the need to fit the approximation at each time step. Function approximation techniques will be revisited in Section 4.5. The most important reason for using the IL algorithm is that we can be aware of the errors e_c and e_v that are introduced through its use, and tune them independently by choosing which numerical integration and function approximation techniques to employ. These choices will be dictated by the dimensionality, the Lévy model, and the payoff function of the problem at hand.

4.2 Software Engineering

We now discuss software design for implementation of Algorithm 4.1.1. In order to separate the implementation details of the specific example that we have chosen from the more generally applicable structure of the program, we used an object-oriented approach. Note that this is not the only possible approach, and the particular design presented here is not the only way to implement Algorithm 4.1.1, but it is apt for a more detailed discussion of model efficiency.

The term *object-oriented* refers to the idea of associating data with the functions that operate upon it in an *object*. The definition of a type of object is called a *class*. For example, we will define a `Surface` class to model value surfaces, $\widehat{v}(n,i,\cdot)$, $i \in \{0,\ldots,N(n)\}$, and a `Distribution` class to provide common functionality associated with the transition probability density function. An object of a particular class is said to be an *instance* of that class. When running a program that imple-

ments Algorithm 4.1.1, we would expect to instantiate the `Surface` class $N(n)+1$ times. A key concept in object-oriented programming is *inheritance*, which allows us to define subclasses of existing classes. A subclass inherits and extends the functionality of its superclass. For instance we will define a `VGDistribution` class as a subclass of the `Distribution` class to provide the transition probability density function for the Variance Gamma process. For a good reference on the subject of object-oriented programming see Gamma, Helm, Johnson, and Vlissides (1995).

We represent the design using a class diagram in Figure 4.1. A class diagram displays the relationship between the different classes used in a program. A triangular arrowhead indicates an inheritance relationship. For example, in Figure 4.1, `WorkerSurface` is a subclass of `Surface`. A solid diamond arrowhead indicates instances of one class contain instances of another and an empty diamond arrowhead shows that an instance of one class contains a reference to the instance of another. For example, a `WorkerSurface` object contains a `Grid` object, and any `Grid` object knows how to find a `Distribution` object. The class boxes in Figure 4.1 show the data and functions that make up each class along with the datatypes of the data, function inputs, and function outputs. Figure 4.1 was generated from the code that was used to obtain the results reported in Section 4.4. Details regarding object creation and destruction have been deleted for clarity. More information on the format and meaning of class diagrams can be found in Gamma et al. (1995, p. 363).

We now explain Figure 4.1. The `Lattice` object serves as both a container for the other objects and to provide an interface that hides the implementation details of the algorithm. A `Lattice` object has a reference to a `Distribution` and a `Payoff`, and contains a `PointSurface` object.

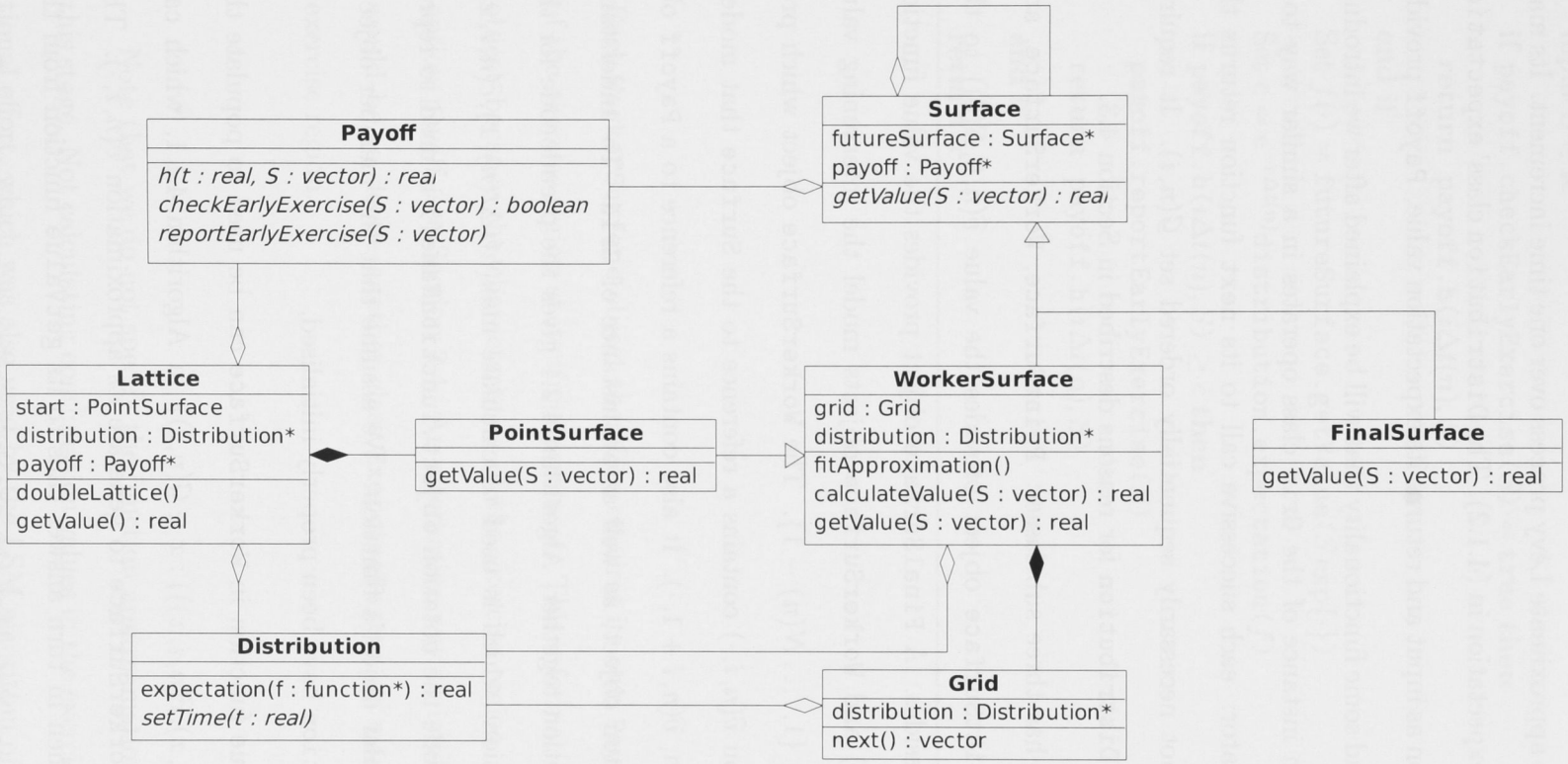

Figure 4.1: Class diagram for a program implementing Algorithm 4.1.1

The `Distribution` class provides functions related to the transition probability density function for the approximate Lévy process over one time increment. Its main role is to calculate the expectation in (4.1.2). The `Distribution` class' `expectation` function takes a function as input and returns its expectation value. `Payoff` provides the payoff function h and some functionality that will be explained after we introduce the `Surface` class. An instance of the `Grid` class operates in a similar way to a random number generator: each successive call to its `next` function returns the next member of the not necessarily sequentially ordered set $\mathcal{G}(n,i)$. It requires information about the `Distribution` for reasons described in Section 4.3.

The `Surface` class has three subclasses: `PointSurface`, `WorkerSurface`, and `FinalSurface`. A `PointSurface` object provides the value $\widehat{v}(n,0,S(0))$ to the `Lattice` in which it resides. A `FinalSurface` object provides the value function $\widehat{v}(n,N(n),x) = h(t,x)$ and `WorkerSurface` objects model the intervening value functions $\widehat{v}(n,i,\cdot)$, $i \in \{1,\ldots,N(n)-1\}$. The `WorkerSurface` object which provides the value function $\widehat{v}(n,i,\cdot)$ contains a reference to the `Surface` that models the next value function, $\widehat{v}(n,i+1,\cdot)$. It also contains a reference to a `Payoff` object and a `Distribution` object, as well as an instance of `Grid`. To understand how these objects function together, Algorithm 4.2.1 gives the pseudocode for the `calculateValue` function, which is used to calculate $\max(h(i\Delta t(n),x),\widehat{c}(n,i,x))$ from (4.1.3). In this code the notation `object.functionName()` is used to represent a call to a particular object's function. We assume that the member objects `payoff` and `distribution` have been properly initialised.

The `calculateValue` function in `WorkerSurface` can be used to populate the list $\{(x,\max(h(i\Delta t(n),x),\widehat{c}(n,i,x))) : x \in \mathcal{G}(n,i)\}$ in Algorithm 4.1.1, which can then be used by the `WorkerSurface` to generate its approximation $\widehat{v}(n,i,\cdot)$. The `WorkerSurface` may then in turn answer calls to its `getValue` function from the $(i-1)$th `WorkerSurface`. Note that the `checkEarlyExercise` and `reportEarlyExercise` functions called in `calculateValue` are provided by a `Payoff` object.

Algorithm 4.2.1 Pseudocode for `WorkerSurface::calculateValue`.

Input $S \in \mathbb{R}^+$

if `payoff.checkEarlyExercise(`S`) = true` **then**

  return `payoff.h(`$i\Delta t(n), S$`)`

end if

Set $f(\cdot)$ = `futureSurface.getValue(`$S\exp\{\cdot\}$`)`

Set $c = \mathrm{e}^{-r\Delta t(n)}$`distribution.expectation(`$f$`)`

if `payoff.h(`$i\Delta t(n), S$`)` $\geq c$ **then**

  `payoff.reportEarlyExercise(`S`)`

  return `payoff.h(`$i\Delta t(n), S$`)`

end if

return c

Each `WorkerSurface` object therefore requires its own `Payoff` object to keep track of the exercise region at its time increment. This information resides with a `Payoff` object because the characteristics of the exercise region are dependent upon the option payoff, see Section 2.4. An alternative way to implement this functionality would be to define another class `ExerciseRegion` to manage information on the exercise region.

Note that we do not specify the function approximation technique employed at this stage. Not calculating continuation values in the exercise region saves computational effort, which was also used by the LSM algorithm of Longstaff and Schwartz (2001), which we reviewed in Section 2.4.

4.3 Pricing an American Put with the Skeleton Approximation Model

In order to be any more specific about the choices that must be made to implement the system described in Section 4.2, we need a particular example, so we take the simplest possible nontrivial example of pricing an American put in one dimension using the skeleton approximation model described in Section 3.3.

We define `PutPayoff` as a subclass of `Payoff`, as illustrated in Figure 4.2. Note that a `PutPayoff` object can replace a `Payoff` object without effecting the other parts of the program. The `h` function in `PutPayoff` is implemented as one would expect. As discussed in Section 2.3, for a single asset American put there exists an exercise boundary $b : [0, T] \mapsto \mathbb{R}^+$ that divides the continuation region from the exercise region, the former being above the boundary and the latter below it. As each `WorkerSurface` possesses its own `Payoff` object, we represent the exercise boundary with a single variable, as seen in Figure 4.2. The `checkEarlyExercise` function therefore returns the Boolean value of (`S` $\leq$ `earlyExerciseBoundary`), and the `reportEarlyExercise` function sets `earlyExerciseBoundary` to the input value `S`, so gaining incrementally better approximations to the real boundary.

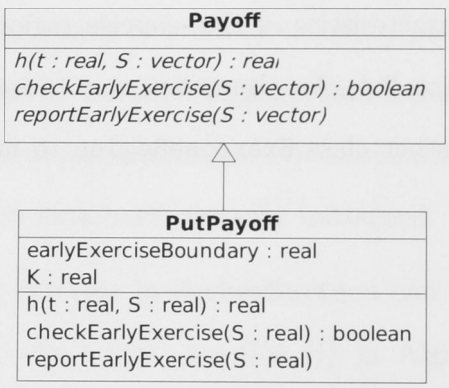

Figure 4.2: Implementation of `Payoff`

We now choose the function approximation scheme for `WorkerSurface`, and for simplicity we use linear interpolation. Using linear interpolation presents a problem

for calculating the value of `getValue` for `S` not in the range $[\min \mathcal{G}(n,i), \max \mathcal{G}(n,i)]$. This can be easily overcome, however, by calling `calculateValue` for any such `S` and adding it as a new point in the grid, thereby adaptively extending it. In this way we avoid the so-called localisation error present in all PIDE approaches that is caused by only solving the PIDE over a finite interval on the price axis. For this approximation scheme, the approximation error defined in (4.1.3) is

$$e_v(n,i,x) \leq \begin{cases} 0, & x \leq \max \mathcal{C}(n,i), \\ \Delta(n,i), & x \in (\max \mathcal{C}(n,i), \min \mathcal{E}(n,i)), \\ \Delta^2(n,i) \max_{y \in \mathcal{E}(n,i)} \left| \frac{\mathrm{d}}{\mathrm{d}y} \widehat{c}(n,i,y) \right|, & x \geq \min \mathcal{E}(n,i), \end{cases}$$

where

$$\begin{aligned} \Delta(n,i) &= \max_{y \in \mathcal{G}(n,i)} \left(\min_{x \in \mathcal{G}(n,i)} (|x-y|) \right), \\ \mathcal{E}(n,i) &= \{x \in \mathcal{G}(n,i) : h(i\Delta t(n), x) < \widehat{c}(n,i,x)\} \\ \text{and} \quad \mathcal{C}(n,i) &= \{x \in \mathcal{G}(n,i) : h(i\Delta t(n), x) \geq \widehat{c}(n,i,x)\}. \end{aligned}$$

Without making any further assumptions about the way in which $\mathcal{G}(n,i)$ is initially generated, which we will discuss in Section 4.4, we assume that $\Delta t(n)$ is proportional to $1/|\mathcal{G}(n,i)|$. In this case we would expect that, as $|\mathcal{G}(n,i)|$ is increased, with all else being held constant, the error in $\widehat{v}(n,0,S(0))$ would reduce at a rate proportional to $|\mathcal{G}(n,i)|^{-2}$, with secondary effects caused by the proximity of the exercise boundary to its closest grid point. These secondary effects are common in lattice models and can cause convergence to be less than smooth. For example, in Lord et al. (2007), provision is made at each time step to move the grid so that a grid point always lies on the exercise boundary. Our grid $\mathcal{G}(n,i)$ could be modified in this way but we avoid this additional complication.

In Figure 4.3, which plots $\widehat{v}(n,0,S(0))$ against $|\mathcal{G}(n,i)|^{-2}$ for a particular numer-

ical experiment, the expected roughly linear relationship is observed. The result of a linear regression on the visible data is also plotted for comparison. Note that, as the value function is always concave for an American put, $e_v(n, i, x)$ is aways positive. For this numerical experiment the parameters have been set so that $N(n) = 32$ and $|e_c(n, i, \cdot)| < 10^{-5}$. The means by which $|e_c(n, i, \cdot)|$ is controlled and the specific underlying Lévy model and option parameters will be described in Section 4.4.

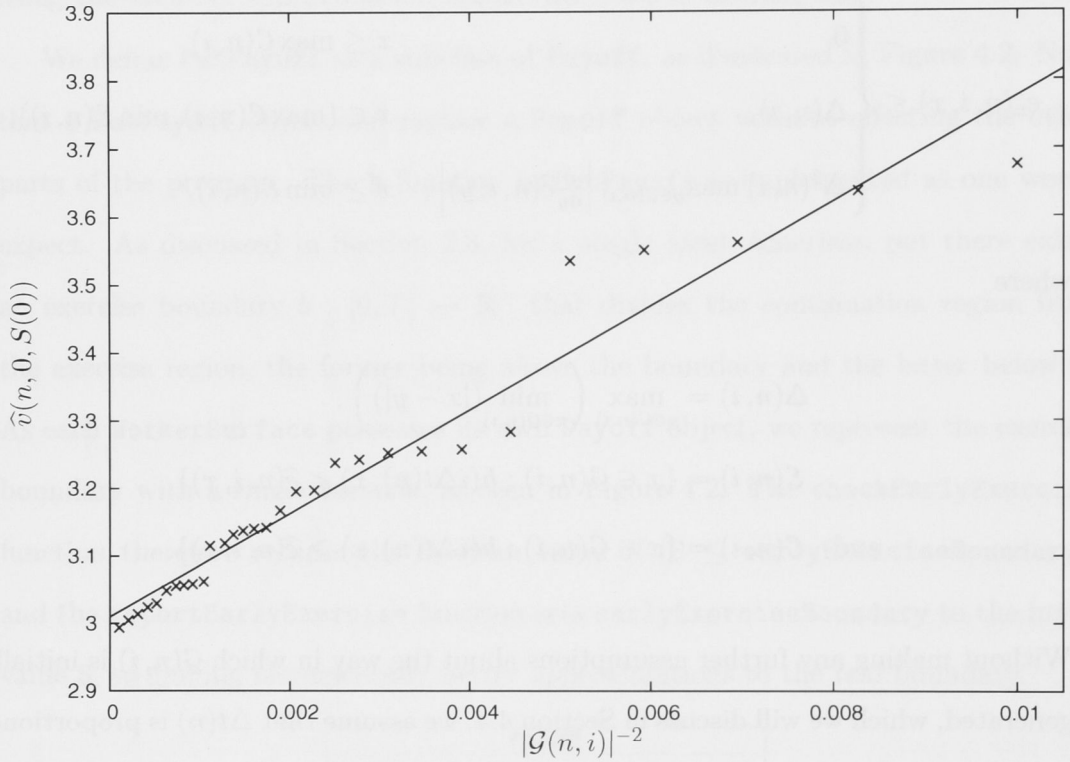

Figure 4.3: Empirical test of the relationship between $\Delta(n, i)$ and $e_v(n, i, \cdot)$, as produced by the example implementation.

C++ was used to write the example implementation, and wherever possible numerical routines from the Gnu Scientific Libary (GSL) were employed. The reference for GSL is Galassi, Davies, Theiler, Gough, Jungman, Booth, and Rossi (2008). We used interpolation routines provided by GSL when implementing `WorkerSurface`. If no assumptions are made about whether $\mathcal{G}(n, i)$ is initially sorted by price, then it must be sorted at a cost of $\mathrm{O}(|\mathcal{G}(n, i)| \log_2 |\mathcal{G}(n, i)|)$ operations (Press et al. 1992, p. 329). Each call to the `getValue` function of a `WorkerSurface` object with an argu-

ment of `S` will then require a search to find the smallest bounding interval for `S` with endpoints drawn from the elements of $\mathcal{G}(n,i)$, unless $S \leq \max \mathcal{C}(n,i)$, in which case the payoff can be returned immediately. Each such search requires $\mathrm{O}(\log_2 |\mathcal{G}(n,i)|)$ operations if we know nothing about the sequence of values that will be passed to `getValue`. However, if successive calls to `getValue` are more likely to be near to one another, then the search can be accelerated to use roughly $\mathrm{O}\left(\log_2^2 |\mathcal{G}(n,i)|\right)$ operations (Press et al. 1992, p. 117). GSL provides accelerated searches.

4.4 The Skeleton Approximation Model with a Variance Gamma Process

The Lévy process chosen for this example is the VG process described in Section 2.2. We reparametrise it slightly by setting $\nu = b^{-1} > 0$, $\theta = \mu \in \mathbb{R}$, and keeping $\sigma > 0$, so that it has the transition probability density function

$$
\begin{aligned}
&f_{\Delta t(n)}\left(x + \delta \Delta t(n)\right) \\
&= \sqrt{\frac{2}{\pi}} \frac{1}{\sigma \nu^{\frac{\Delta t(n)}{\nu}} \Gamma\left(\frac{\Delta t(n)}{\nu}\right)} \left(\frac{x^2}{\theta^2 + \frac{2\sigma^2}{\nu}}\right)^{\frac{\Delta t(n)}{\nu} - \frac{1}{4}} \exp\left\{\frac{\theta x}{\sigma^2}\right\} \mathrm{K}_{\frac{\Delta t(n)}{\nu} - \frac{1}{2}}\left(\left(\frac{\theta^2}{\sigma^2} + \frac{2}{\nu}\right)^{\frac{1}{2}} \frac{|x|}{\sigma}\right) \\
&= C(n) |x|^{D(n)} \mathrm{e}^{Ax} \, \mathrm{K}_{D(n)}(B|x|), \qquad (4.4.1)
\end{aligned}
$$

where K_α is the modified Bessel function of the second kind (Abramowitz and Stegun 1964, p. 364) and

$$
\begin{aligned}
\delta &= r + \frac{1}{\nu} \ln\left(1 - \theta\nu - \frac{\sigma^2 \nu}{2}\right), \\
A &= \frac{\theta}{\sigma^2}, \\
B &= \frac{1}{\sigma}\left(\frac{\theta^2}{\sigma^2} + \frac{2}{\nu}\right)^{\frac{1}{2}},
\end{aligned}
$$

$$D(n) = \frac{\Delta t(n)}{\nu} - \frac{1}{2},$$
$$\text{and} \quad C(n) = \sqrt{\frac{2}{\pi}} \left(B^{D(n)} \left(\sigma^2 \nu\right)^{\frac{\Delta t(n)}{\nu}} \Gamma\left(\frac{\Delta t(n)}{\nu}\right)\right)^{-1}.$$

Note the drift correction term δ which ensures that $\mathrm{E}\, \mathrm{e}^{L(n,\Delta t(n))} = \mathrm{e}^{r\Delta t(n)}$, see Section 2.1.

Remark 4.4.1. *When implementing* (4.4.1)*, roundoff error (Press et al. 1992, p. 28) can be brought within acceptable levels by evaluating*

$$\exp\left\{\ln(\mathrm{K}_{D(n)}(B|x|)) + D(n)\ln|x| + Ax + \ln C(n))\right\},$$

where $\ln(\mathrm{K}_\alpha(\cdot))$ *is evaluated in a single function call to the appropriate GSL function.*

We implement `VGDistribution`, a subclass of `Distribution`, to evaluate the expectation

$$\mathrm{E}\, g(L^{\mathrm{D}}(n, \Delta t(n))) = \int_{-\infty}^{\infty} g(x) f_{\Delta t(n)}(x)\, \mathrm{d}x, \tag{4.4.2}$$

where $L^{\mathrm{D}}(n,t)$ is as defined in (3.3.1), for an arbitrary function $g : \mathbb{R} \mapsto \mathbb{R}^+$ using numerical integration procedures supplied by GSL. If $D(n) \geq 0$, that is $\Delta t(n) \geq \nu/2$, this is relatively straightforward, and we can apply the QAGI numerical integration routine (Galassi et al. 2008, p. 179). As for all GSL integration routines, one of QAGI's inputs is requested accuracy. This input can be in either relative or absolute terms. As the absolute error of the integral corresponds to our definition of $e_c(n, i, \cdot)$ in (4.1.2), we request a constant absolute error in our calls to GSL integration routines.

QAGI uses the transformation $x = (1-y)/y$ to evaluate

$$\begin{aligned}&\mathrm{E}\, g(L^{\mathrm{D}}(n, \Delta t(n)))\\ &= \int_0^1 \frac{1}{y^2}\left(g\left(\frac{1-y}{y}\right) f_{\Delta t(n)}\left(\frac{1-y}{y}\right) + g\left(\frac{y-1}{y}\right) f_{\Delta t(n)}\left(\frac{y-1}{y}\right)\right) \mathrm{d}y,\end{aligned}$$

so that the integration interval is finite. See Press et al. (1992, Section 4.4) for details on numerically calculating improper integrals. QAGI then calls another GSL integration function, QAGS, to perform the integration. Like many GSL integration algorithms, QAGS is an adaptive integration technique which bisects its integration range, then applies two integration rules to each subinterval. This allows it to estimate the integration error on each subinterval, and select the interval with the greatest error for further bisection. Bisection of the interval with the greatest estimated error is repeated until either an estimated error criterion is met or a maximum number of subdivisions is reached. If the latter condition terminates execution, the algorithm is considered to not have converged. QAGS uses Gauss-Kronrod quadrature rules (Press et al. 1992, p. 160) on each subinterval, and is designed to speed convergence of integration of integrands with singularities using a special extrapolation technique (Galassi et al. 2008, p. 179).

If $D(n) < 0$, that is $\Delta t(n) < \nu/2$, then the integration in (4.4.2) is complicated by the presence of an integrable singularity in $f_{\Delta t(n)}(y)$ at $y = \delta \Delta t(n)$. To see how this occurs, we apply an approximation to K_α given in Abramowitz and Stegun (1964, p. 375) to obtain

$$f_{\Delta t(n)}\left(x + \delta \Delta t(n)\right) \approx C(n) \frac{\Gamma(|D|)}{2} \left(\frac{2}{B}\right)^{|D|} |x|^{D(n)-|D(n)|},$$

for x near zero. There are a number of routines supplied by GSL for integrating functions with integrable singularities. The QAGS procedure, mentioned above, will integrate integrands with singularities. However, QAGP (Galassi et al. 2008, p. 179)

will converge faster than QAGS if the singular points are known prior to integration. QAGP uses QAGS in its implementation. We can therefore use QAGS to calculate the integral

$$\int_{-1}^{1} g(x+\delta\Delta t(n)) f_{\Delta t(n)}(x+\delta\Delta t(n))\,\mathrm{d}x,$$

by supplying it with the known singular point of $x = 0$. The two tail integrals,

$$\int_{-\infty}^{-1} g(x+\delta\Delta t(n)) f_{\Delta t(n)}(x+\delta\Delta t(n))\,\mathrm{d}x$$

$$\text{and} \quad \int_{1}^{\infty} g(x+\delta\Delta t(n)) f_{\Delta t(n)}(x+\delta\Delta t(n))\,\mathrm{d}x,$$

can then be calculated using QAGIL and QAGIU, which function in a similar fashion to QAGI but for semi-infinite integrals. If an integrand takes the form

$$\int_{a}^{b} (x-a)^{\alpha}(b-x)^{\beta} f(x)\,\mathrm{d}x, \quad \alpha,\beta > -1,$$

then the QAWS routine provides another alternative (Galassi et al. 2008, p. 181). QAWS also uses an adaptive approach, applying Gauss-Kronrod quadrature away from the end points, and a modified Clenshaw-Curtis quadrature rule (Press et al. 1992, p. 196) at the end points. Clenshaw-Curtis quadrature functions by integrating a Chebyshev approximation (Press et al. 1992, Sections 5.8–10) to the integrand. Thus we can use QAWS to evaluate

$$\int_{-1}^{0} |x|^{2D(n)} g(x+\delta\Delta t(n)) C(n) |x|^{-D(n)} \mathrm{e}^{Ax}\, \mathrm{K}_{D(n)}(B|x|)\,\mathrm{d}x$$

$$\text{and} \quad \int_{0}^{1} x^{2D(n)} g(x+\delta\Delta t(n)) C(n) x^{-D(n)} \mathrm{e}^{Ax}\, \mathrm{K}_{D(n)}(Bx)\,\mathrm{d}x,$$

in conjunction with the above tail integrals. Through experimentation it was discovered that the best approach to calculating (4.4.2) was to use both the QAGP and QAWS algorithms. That is, apply the QAGP routine, and if it fails to converge,

then take the output of the QAWS function. This provided more stable results than using either technique in isolation.

An empirical test was used to explore the efficiency of the above integration methods, where the total execution time of the pricing program was recorded as more stringent absolute error criteria were given to the GSL integration routines. In this test the parameters were set at $|\mathcal{G}(n,i)| = 128$ and $N(n) = 64$. The results are shown in Figure 4.4, which maps the total execution time against in seconds against the base ten logarithm of the requested absolute error. A linear regression to the data in Figure 4.4 is plotted for comparison. The program was executed on a server with four dual-core AMD Opteron 275 processors running at 2.2 GHz and 4 GB of shared memory, although the program was not multi-threaded. The pricing parameters for this test were the same as those used for the tests that generated Figure 4.5, and are given below.

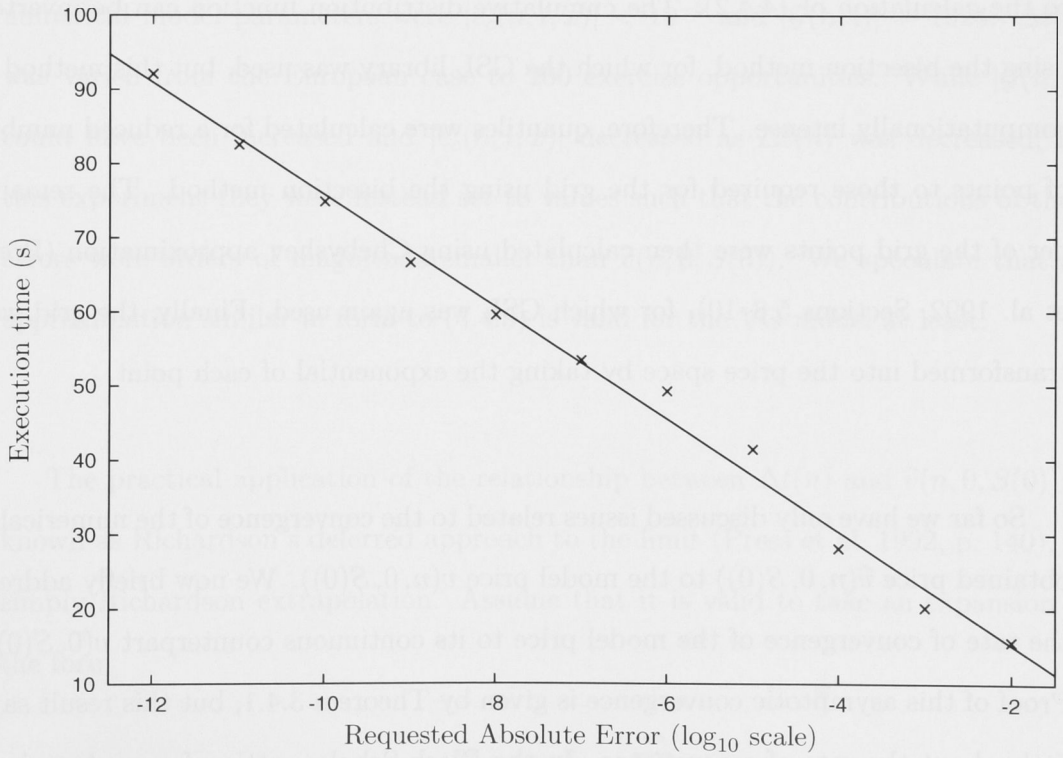

Figure 4.4: Execution time as a function of requested absolute integration error.

We now give more specific information about how the grid $\mathcal{G}(n,i)$ was generated for our example. In Kargin (2005), the author employed the heuristic of generating a quasi-random sequence in the $[0,1]^d$ hypercube, then applying an inverse mapping of the cumulative density function of the multidimensional Normal distribution. Quasi-random sequences are similar to pseudo-random sequences, but rather than attempting to produce independent samples, they attempt to produce samples that, in a well defined way, maximally avoid all of the preceding sample points. As they fill a space more efficiently than pseudo-random points, they can be used, for example, to accelerate Monte Carlo integration methods. Quasi-random sequences are reviewed in Press et al. (1992, Section 7.7). We use the same technique as Kargin for our example. This requires the generation of a quasi-random sequence, for which the GSL Sobol sequence generator (Press et al. 1992, p. 310) was used. Next, calculation of the VG transition cumulative distribution function requires the integration of (4.4.1). This can be accomplished in an exactly analogous manner to the calculation of (4.4.2). The cumulative distribution function can be inverted using the bisection method, for which the GSL library was used, but this method is computationally intense. Therefore, quantiles were calculated for a reduced number of points to those required for the grid using the bisection method. The remainder of the grid points were then calculated using Chebyshev approximation (Press et al. 1992, Sections 5.8–10), for which GSL was again used. Finally, the grid was transformed into the price space by taking the exponential of each point.

So far we have only discussed issues related to the convergence of the numerically obtained price $\widehat{v}(n,0,S(0))$ to the model price $v(n,0,S(0))$. We now briefly address the rate of convergence of the model price to its continuous counterpart $v(0,S(0))$. Proof of this asymptotic convergence is given by Theorem 3.4.1, but this result says little about the rate of convergence. In the Black-Scholes setting for an American put on a single variable, Howison (2007) uses perturbation theory to derive the outer

expansion

$$v(n,t,S(t)) \approx v(t,S(t)) - \Delta t(n) v_2(t,S(t)) + \mathrm{O}\left(\Delta t(n)^{\frac{3}{2}}\right), \tag{4.4.3}$$

which is valid in the region $S(t)/b(t) - 1 \gg \mathrm{O}\left(\Delta t(n)^{\frac{1}{2}}\right)$, where $v_2 : [0,T] \times \mathbb{R}^+ \mapsto \mathbb{R}^+$ and b is the exercise boundary. An inner expansion which is more accurate for $S(t)$ close to the exercise boundary is also provided.

The rate of convergence of $\widehat{v}(n,0,S(0))$ to $v(n,0,S(0)$ for general Lévy models has had little research. We present some empirical results in Figure 4.5, in which $\hat{v}(n,0,S(0))$ is mapped against $\Delta t(n)$ using a VG model. Taking (4.4.3) as motivation, a curve of the form $a + b\Delta t(n) + c\Delta t(n)^{\frac{3}{2}}$ is fitted to the results. The VG model had parameters $\theta = -0.14$, $\sigma = 0.12$, and $\nu = 0.2$. The option parameters were starting price $S(0) = 100$, strike price $K = 100$, and maturity $T = 1$. The numerical model parameters were $|e_c(n,i,x)| < 10^{-5}$ and $|\mathcal{G}(n,i)| = 1024$. $\Delta t(n)$ was varied from the European case to 200 exercise opportunities. While $|\mathcal{G}(n,i)|$ could have been increased and $|e_c(n,i,x)|$ decreased as $\Delta t(n)$ was decreased, for this experiment they were instead set to values such that the contributions of their errors were orders of magnitude smaller than $\widehat{v}(n,0,S(0))$. We speculate that an approximation similar in form to (4.4.3) is valid for the VG model at least.

The practical application of the relationship between $\Delta t(n)$ and $\widehat{v}(n,0,S(0))$ is known as Richardson's deferred approach to the limit (Press et al. 1992, p. 140), or simply Richardson extrapolation. Assume that it is valid to take an expansion of the form

$$v(n,0,S(0)) = v(0,S(0)) + \sum_{k=1}^{\infty} a_k \Delta t(n)^{\gamma_k},$$

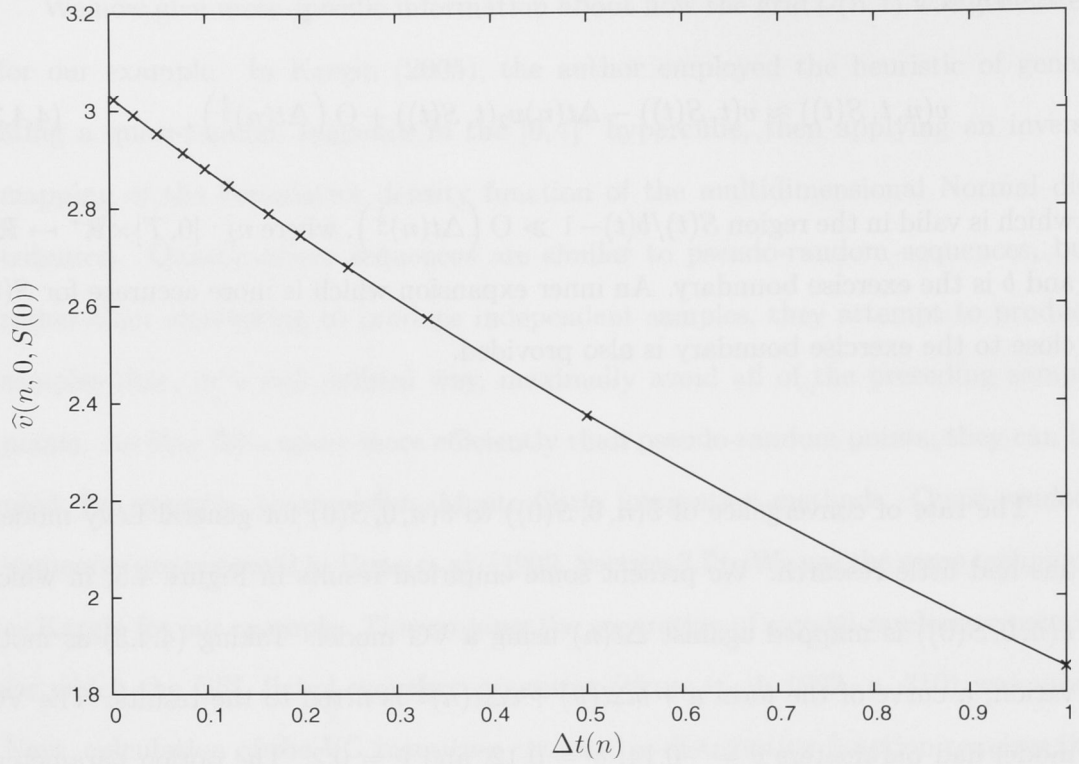

Figure 4.5: Convergence of Bermudan to American prices for the VG model.

where $0 < \gamma_k < \gamma_{k+1}$ and $a_k \in \mathbb{R}$. If the γ_k are known, an expression of the form

$$\widehat{f}(\Delta t(n)) = \widehat{v}(0, S(0)) + \sum_{k=1}^{K-1} \widehat{a}_k \Delta t(n)^{\gamma_k},$$

can be fitted to the data $\{(\Delta t(n_k), v(n_k, 0, S(0)))\}_{k\in\{1,\ldots,K\}}$ for some sample $\{n_k\}_{k\in\{1,\ldots,K\}}$. $\widehat{f}(\Delta t(n))$ can then be evaluated at $\Delta t(n) = 0$ to obtain an estimate for $v(0, S(0))$ with error $\mathrm{O}\left(\Delta t(n_1)^{\gamma_K}\right)$. This technique accelerates convergence of the estimates for $v(0, S(0))$ and gives an error estimate. Lord et al. (2007) mentions that the choice of $\gamma_k = k$ appears to work well for VG and CGMY processes. Richardson extrapolation has also been used to speed convergence of lattice models in Chang, Chung, and Stapleton (2007), Geske and Johnson (1984), and others. Based on the results shown in Figure 4.5, Richardson extrapolation should be implemented for any application of the algorithm presented in Section 4.1 where speed of execution is a concern.

4.5 Considerations for Higher Dimensional Problems

In this chapter we have presented an example implementation as an aid to explaining the issues and trade-offs associated with implementing a pricing model based on the theory developed in Chapter 3. An implementation of Algorithm 4.1.1 for pricing an American rainbow option would require two elements: a fast method of adaptive interpolation to find $\widehat{v}(n, i, \cdot)$ in (4.1.3) and a matched method of efficiently calculating the integral implicit in the calculation of $\widehat{c}(n, i, \cdot)$ in (4.1.2). This section discusses some lines of enquiry that could yield these elements.

Firstly, however, it should be mentioned that the techniques we used for generating the grid $\mathcal{G}(n, i)$ in Section 4.4 can be generalised to higher dimensions, but for a general transition probability density function this is difficult. For the skeleton approximation model, a better approach would be to simulate paths, using either pseudo- or quasi-random number generators, of the approximation model and using their values at $i\Delta t(n)$, $i \in \{1, \ldots, N\}$ as grid points. Good algorithms for simulating many Lévy models and their use in multiple dimensions is discussed in Cont and Tankov (2004, Chapter 6 and Section 11.5).

Returning to the relationship between interpolation and integration methods, the pricing model implemented in Kargin (2005), which was reviewed in Section 2.4, succeeds because the approximating function $\widehat{v}(t, x)$ given in (2.4.3) is a sum of Gaussian functions, which is then integrated against another Gaussian function, the transition probability density function, to calculate the continuation value. Kargin shows that the approximation function can be found efficiently, and the integration is fast because it reduces to a sum of known integrals. Thus it may be possible to find methods of approximation that actually reduce the complexity of the multidimensional numerical integrals required in (4.1.2) for some Lévy models.

It is not difficult to conceive of a situation where the interpolation method could

aid the evaluation of (4.1.2) in one dimension. If we consider the use of linear interpolation, as described in Section 4.3, with the first jump approximation model developed in Section 3.2 for the VG model, then by (2.2.1) and (3.2.4), the transition probability density function becomes

$$\begin{aligned} & f_{\Delta t(n)}(x+\delta(n))\,\mathrm{d}x \\ &= \frac{1-\mathrm{e}^{-\Delta t(n)\Pi\{\mathcal{J}(n)\}}}{\Pi\{\mathcal{J}(n)\}}\frac{b}{|x|}\exp\left\{\frac{\mu x}{\sigma^2}-\left(\frac{\mu^2}{\sigma^2}+2b\right)^{\frac{1}{2}}\frac{|x|}{\sigma}\right\}\mathbf{1}_{\mathcal{J}(n)}(x)\,\mathrm{d}x \\ &\quad + \mathrm{e}^{-\Delta t(n)\Pi\{\mathcal{J}(n)\}}\delta_0\,(\mathrm{d}x)\,. \end{aligned}$$

If we specify the grid $\mathcal{G}(n,i) = \{g(n,i,k)\}_{k\in\{0,\ldots,K(n,i)\}}$, $g(n,i,k) < g(n,i,k+1)$, then (4.1.2) becomes

$$\widehat{c}(n,i,x) = \mathrm{e}^{-r\Delta t(n)}\sum_{k=1}^{K(n,i)}\int_{g(n,i,k-1)}^{g(n,i,k)}(a(n,i,k)x\mathrm{e}^y + b(n,i,k))f_{\Delta t(n)}(y)\,\mathrm{d}y,$$

where $a(n,i,k)$ and $b(n,i,k)$ are coefficients of the linear interpolation in the region $[g(n,i,k-1), g(n,i,k))$. This reduces to integrals of the form

$$\int_a^b \mathrm{e}^x\frac{\mathrm{e}^{-Ax}}{x}\,\mathrm{d}x = \mathrm{E}_1(a(1-A)) - \mathrm{E}_1(b(1-A))$$

$$\text{and}\quad \int_a^b \frac{\mathrm{e}^{-Ax}}{x}\,\mathrm{d}x = \mathrm{E}_1(aA) - \mathrm{E}_1(bA),$$

where $\mathrm{E}_1(x) = \int_1^\infty t^{-1}\mathrm{e}^{-xt}\,\mathrm{d}t$ is the exponential integral (Press et al. 1992, p. 222), a standard function that can be evaluated efficiently using, for example, GSL. However, while this method would work with ordinary linear interpolation, it could be made faster still if the number of linear segments $[g(n,i,k-1), g(n,i,k))$ could be minimised efficiently, say with a good adaptive interpolation algorithm.

An example of an adaptive algorithm that yields piecewise linear approximations is the nonlinear approximation as given in (2.4.3) in conjunction with the hinging

hyperplanes of Breiman (1993), although this function approximation algorithm is better suited to higher dimensional problems. Hinging hyperplane functions are, as the name would suggest, two half planes joined along their intersection. Another algorithm that would be applicable to this task, and is also suitable for use in higher dimensions, is the multivariate adaptive regression splines (MARS) algorithm (Friedman 1991a,b). This nonparametric, adaptive function approximation algorithm does not appear to have been adopted in the finance literature, but could be useful for the type of multivariate function approximation we require for higher dimensional interpolative lattice algorithms.

The use of nonlinear function approximation methods such as these that yield very simple representations of their target functions hold much potential for simplifying the otherwise onerous multidimensional integrals implicit in (4.1.2). This is particularly true for low-dimensional American rainbows where, as in Section 2.2.2, closed form solutions for Lévy measures and transition probability density functions can be discovered.

Chapter 5

Conclusion

In Part I there are two main contributions. The first is the generalisation of the multinomial model of Maller, Solomon, and Szimayer (2006) with associated convergence results to the case of rainbow options with an arbitrary number of underlying assets. This is embodied in Theorem 3.4.2. Given the inherent difficulties of implementing any numerical method for evaluating integrals in multiple dimensions (Press, Teukolsky, Vetterling, and Flannery 1992, p. 161), we have attempted to state the multidimensional first jump approximation model in a way that is conducive to fast implementations.

The second contribution is a proof that in a Lévy model setting, the price of a Bermudan option converges to that of the corresponding American option as the number of exercise opportunities increases. Theorem 3.4.1, which gives this result, serves as the archetypal example application of Theorem 3.3.1 with Theorem 5 of Coquet and Toldo (2007). This approach allows us to prove the convergence of other multinomial models, as seen in the example shown in Section 3.5. We also give a generalisation of Theorem 3.4.1 to multiple dimensions.

There are several secondary results. Closed form expressions are provided for the Lévy measures of the two dimensional Variance Gamma process with a common subordinator, the $\alpha\beta$-Variance Gamma process of Semeraro (2008), and the

two dimensional Normal Inverse Gaussian process with a common subordinator. Analytic expressions are also given for the transition probability density function of the two common subordinator process. This opens these models for use pricing two-colour American rainbow options with either the skeleton or first jump approximation models. Further, an example implementation was given to demonstrate some of the practical ramifications of the theoretical results.

There is much potential for future work. Some of the opportunities for extending the theoretical developments in Chapter 3 were discussed in Section 3.6. Generalisation of Theorem 5 in Coquet and Toldo (2007) for multidimensional stochastic processes would precipitate two convergence results similar in form to Theorem 3.4.1 for American rainbow options under the first jump and skeleton approximation models.

On a more practical note, the flexibility of the skeleton and first jump approximation models allows for a multitude of possible implementations of lattice models for pricing American rainbow options using Lévy processes. As explained in Section 4.5, there is considerable scope for creative solutions in the face of the curse of dimensionality, which could open several interesting lines of enquiry.

Part II

A Generalised Skewness Statistic

Chapter 6

A Generalised Skewness Statistic

6.1 Introduction

In the fields of asset allocation and derivatives pricing, the analytical tractability of the normal distribution has made possible the significant technological breakthroughs that have been seen in those areas over the last 40 years. Many of the models still used in industry start with the assumption that the returns of asset prices are normally distributed. For instance, for derivatives pricing, the standard model of Geometric Brownian Motion (GBM) for asset prices assumes that log returns are normally distributed and independent (Black and Scholes 1973). It is widely accepted, however, that the daily log returns of stock prices are not best fitted by normal distributions, and that financial time series exhibit certain *stylised features*. See Rydberg (2000) for an excellent survey of this topic. Three such features that are of immediate interest are the leptokurtosis and non-zero skewness of distributions, and serial dependence in time series of daily log returns. Alternative models not predicated on normally distributed log returns have been developed in response. Part I concentrated on exponential Lévy models, which, as was mentioned in Chapter 2, form a much wider class of models that contains GBM.

Some initial motivation for the work presented in this chapter came from the

proposal of a particular alternative stock price model, the Fractal Activity Time Geometric Brownian Motion (FATGBM) model of Heyde (1999), which admits leptokurtosis and strong dependence but does not allow non-zero skewness. A series of daily log returns $\{X_i\}_{i\in\{1,\ldots,n\}}$, $n \in \mathbb{N}$, derived from a FATGBM process satisfies (Heyde and Leonenko 2005):

$$X_i = \mu + \sigma[W(T_i) - W(T_{i-1})],$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are constants. The subordinator $\{T_i\}_{i\in\{1,\ldots,n\}}$ is an asymptotically self-similar stochastic process such that, for large t,

$$T_{ct} - ct \overset{d}{\cong} c^H(T_t - t),$$

where c and H are constants such that $c > 0$ and $0 < H < 1$. $\{W(t)\}_{t\in[0,T_n]}$ is a standard Wiener process that is independent of the subordinator. Note that strong serial dependence between returns can be observed via significant autocorrelation of absolute returns for large lags (Heyde 1999; Rydberg 2000). When assessing the suitability of FATGBM as a risky asset price model, it would therefore be useful to be able to make statements about the significance of the skewness of a sample of daily log returns, without making any assumptions about the degree of kurtosis of its distribution or the independence of its samples.

Leptokurtosis may be measured in terms of the tail weight index of a distribution, which is the index of the power law that best approximates the tail of that distribution. The tail weight index of log return distributions has been estimated to take values between 3 and 6 in Jansen and de Vries (1991), Hurst and Platen (1997), Hurst, Platen, and Rachev (1997), Daníelsson and de Vries (1997, 1998), Bingham, Kiesel, and Schmidt (2003), and Platen and Sidorowicz (2008). As we will demonstrate in the next section, if a sample distribution exhibits leptokurtosis to the degree suggested by these studies, it raises serious questions about whether

we can say anything about the significance of the usual skewness statistic, which is based on the sample third moment.

This chapter is to proposes a class of skewness statistics that accommodates the above-mentioned stylised features of stock returns while not divorcing the statistics too much from the standard skewness statistic. Assumptions of independence or normality, or of any specific distribution, are dropped in favour of assuming only that mean corrected returns are stationary, ergodic martingale differences. Moment assumptions are also significantly weakened. As we observed in the Introduction, assuming that drift corrected prices form a martingale is one way of stating the efficient market hypothesis.

6.2 Theoretical Results

The standard definition of the skewness of a random variable X is

$$\gamma = \frac{\mathrm{E}\,(X-\mu)^3}{\left[\mathrm{E}\,(X-\mu)^2\right]^{\frac{3}{2}}},$$

where $\mu = \mathrm{E}\,X$ and the moments are assumed to be finite. Given a sample $\{X_i\}_{i\in\{1,\ldots,n\}}$, this can be estimated by the statistic

$$\gamma_n = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}_n\right)^3}{\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}_n\right)^2\right]^{\frac{3}{2}}}, \tag{6.2.1}$$

where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. The challenge in this standard situation is to determine whether γ is significantly different to zero. If the $\{X_i\}_{i\in\{1,\ldots,n\}}$ are independent and normally distributed, then under the null hypothesis $\gamma = 0$, we have that $\sqrt{n}\gamma_n \xrightarrow{\mathrm{D}} \mathrm{N}(0,6)$ as $n \to \infty$ (Kendall, Stuart, and Ord 1994, p. 101). If the assumption of normality is replaced by the assumption that $\mathrm{E}\,|X|^6 < \infty$, then more generally we

have that $\sqrt{n}\hat{\gamma}_n \xrightarrow{\mathrm{D}} \mathrm{N}(0, \varsigma^2)$, where

$$\varsigma^2 = \frac{\mathrm{E}(X-\mu)^6}{[\mathrm{E}(X-\mu)^2]^3} - 6\frac{\mathrm{E}(X-\mu)^4}{[\mathrm{E}(X-\mu)^2]^2} + 9. \tag{6.2.2}$$

This follows from Equations (6.2.5) and (6.2.7) below. As an example, if we were to assume that the $\{X_i\}_{i\in\{1,\ldots,n\}}$ were drawn from a t-distribution with ν degrees of freedom, then standard calculations give that

$$\varsigma^2 = 15\frac{(\nu-2)^2}{(\nu-6)(\nu-4)} - 18\frac{\nu-2}{\nu-4} + 9.$$

Figure 6.1 plots values of ς for t-distributions with $6 < \nu \leq 12$, compared with $\varsigma = \sqrt{6}$ for a normal distribution. This comparison illustrates that if we assume normality when the sample is even mildly leptokurtic that the rejection region for a test based on the $\gamma = 0$ hypothesis will be grossly overestimated.

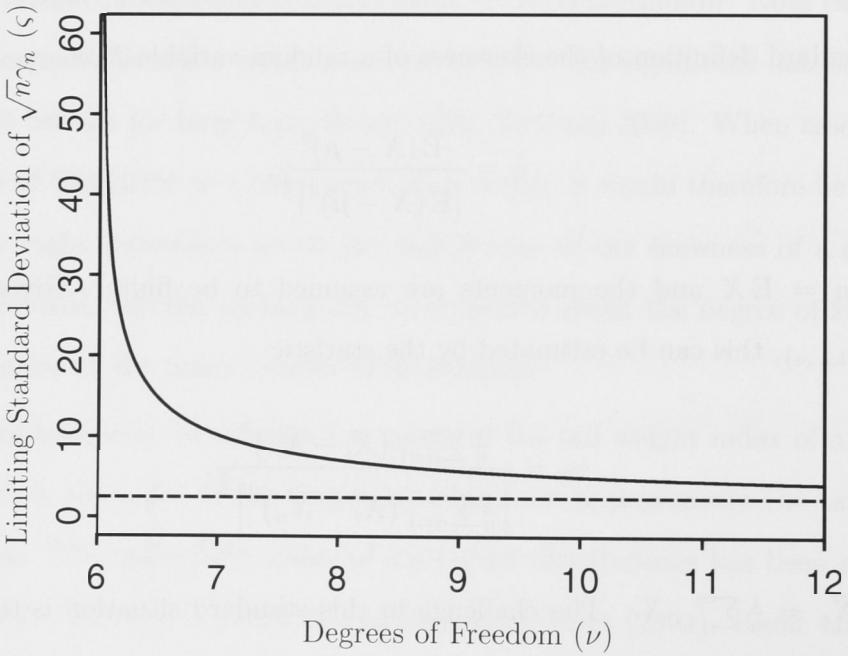

Figure 6.1: ς if log returns follow a Student's t. Dashed line shows $\varsigma = \sqrt{6}$.

If it is suspected that $\mathrm{E}|X|^6$ is not finite, we are unable to make any claim about the significance of γ_n in this way. As described in Section 6.1, the existence

of leptokurtosis in daily stock log returns is widely accepted. Specifically, there is much evidence that log returns can be fitted well by t-distributions (see, for instance, Hurst and Platen 1997; Hurst et al. 1997; Heyde and Liu 2001; Bingham et al. 2003), typically with degrees of freedom $3 < \nu < 5$. In this case the distribution of $\sqrt{n}\gamma_n$ does not tend to normality and a new approach is required.

The solution proposed here is to define a class of modified statistics,

$$g(\beta) = \frac{\mathrm{E}\,\mathrm{sgn}(X-\mu)|X-\mu|^{\beta}}{\left[\mathrm{E}\,(X-\mu)^2\right]^{\frac{\beta}{2}}},$$

depending on a parameter $\beta > 0$, for which we suggest the natural estimator

$$g_n(\beta) = \frac{\frac{1}{n}\sum_{i=1}^{n} \mathrm{sgn}\left(X_i - \overline{X}_n\right)\left|X_i - \overline{X}_n\right|^{\beta}}{\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2\right]^{\frac{\beta}{2}}}. \tag{6.2.3}$$

Note that $\beta = 3$ recovers the "usual" statistic (6.2.1), while $g_n(1) \equiv 0$.

Two advantages arise from this modification. The first is that $g_n(\beta)$ may have reasonable behaviour even when the sixth moment is infinite, which is often the case in financial data, as we observed above. The second is the extra flexibility allowed by the introduction of the parameter β, which allows a wider range of assessment of skewness, as we illustrate in Chapter 7.

To be practically useful, we need information on the distribution of $g_n(\beta)$. We show that, under some very general conditions, $\sqrt{n}g_n(\beta)$ tends in distribution to normality. This provides the machinery for testing the null hypothesis of symmetry of the sample distribution, that is, whether, for some choice of β, $H_0 : g(\beta) = 0$, versus $H_A : g(\beta) \neq 0$.

When $\mu = 0$, $g(\beta)$ simplifies considerably, and an alternative means of estimating the significance of $g_n(\beta)$ for stock log returns is first to test the hypothesis that $\mu = 0$. If it is accepted, the simpler version can be used. A description of this method is included in Appendix 6.5. However we do not assume that $\mu = 0$ in this section.

We proceed to investigate the asymptotic distribution of $g_n(\beta)$ in stages. The following gives the limiting distribution of the numerator of $\sqrt{n}g_n(\beta)$.

Theorem 6.2.1. *Assume we are given a sample of random variables* $\{X_i\}_{i\in\{1,\dots,n\}}$ *defined on a filtered probability space* $(\Omega, \mathcal{F}_n, \mathrm{P})$ *with* $\overline{X}_n = \sum_{i=1}^n X_i/\,n$. *Take* $\beta > 0$ *and assume that*

(A1) $\mathrm{E}\,X^{2\vee 2\beta} < \infty$

(A2) and X *has probability density function* $f_X(x)$ *satisfying* $\sup_{x\in\mathbb{R}} f_X(x) < \infty$,

and let $\mu = \mathrm{E}\,X$. *Assume further that either*

(A3a) $\beta > 1$ *and*

(i) $\{X_i - \mu\}_{i\in\{1,\dots,n\}}$ *are stationary ergodic martingale differences, having*

(ii) $\mathrm{E}\left(\mathrm{sgn}(X_i-\mu)|X_i-\mu|^\beta \middle|\, \mathcal{F}_{i-1}\right) = 0$ *for all* $i \in \{1,\dots,n\}$;

or

(A3b) $0 < \beta < 1$ *and*

(i) $\{X_i\}$ *are iid, with*

(ii) $\mathrm{E}\,\mathrm{sgn}(X-\mu)|X-\mu|^\beta = 0$.

Then the following holds as $n \to \infty$:

$$T_n(\beta) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{sgn}(X_i - \overline{X}_n) \left|X_i - \overline{X}_n\right|^\beta \xrightarrow{\mathrm{D}} \mathrm{N}\left(0, \tau^2(\beta)\right), \tag{6.2.4}$$

where, with $W = X - \mu$,

$$\tau^2(\beta) = \mathrm{E}\,|W|^{2\beta} - 2\beta\,\mathrm{E}\,|W|^{\beta-1}\,\mathrm{E}\,|W|^{\beta+1} + \beta^2\left[\mathrm{E}\,|W|^{\beta-1}\right]^2 \mathrm{E}\,W^2. \tag{6.2.5}$$

The proof for all theorems in this section can be found in Appendix 6.4.

Remark 6.2.1. *Of course* $T_n(1) \equiv 0$, *but (6.2.4) is still true in a degenerate sense then since also* $\tau^2(1) \equiv 0$.

In (6.2.5), $\tau^2(\beta)$ is given in terms of population moments which must be estimated from the sample. The next theorem shows that centralised sample fractional order moments approach their theoretical values for large samples. It is the next step in allowing us to estimate the denominator of $\sqrt{n}g_n(\beta)$. We have to slightly strengthen assumption (A1). Let $\lceil x \rceil$ denote the smallest integer larger than a number x.

Theorem 6.2.2. *Assume that* $\beta > 0$, *that* $\mathrm{E}\,|X|^{2\vee\lceil 2\beta\rceil} < \infty$, *that (A2) holds, and that* $\{X_i - \mu\}_{i\in\{1,\ldots,n\}}$ *are stationary and ergodic. Then, as* $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} \left|X_i - \overline{X}_n\right|^{\alpha} \xrightarrow{\text{a.s.}} \mathrm{E}\,|X-\mu|^{\alpha}, \tag{6.2.6}$$

for any $0 < \alpha \leq 2 \vee \lceil 2\beta \rceil$.

Under the assumptions of Theorems 6.2.1 and 6.2.2, we have that $\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \xrightarrow{\text{a.s.}} \mathrm{E}(X-\mu)^2$ as $n \to \infty$, so we obtain from (6.2.3), (6.2.4), and (6.2.6) that $g_n(\beta) \xrightarrow{\text{D}} \mathrm{N}\left(0, \sigma^2(\beta)\right)$ as $n \to \infty$, where

$$\sigma^2(\beta) = \frac{\tau^2(\beta)}{[\mathrm{E}(X-\mu)^2]^{\beta}}. \tag{6.2.7}$$

In our empirical study we will keep $\beta > 1$. Thus (A3b) is not relevant in the current context, but is included for completeness. Note that Theorem 6.2.2 does not allow us to estimate $\tau^2(\beta)$ in the case $0 < \beta < 1$, as that would require the estimation of $\mathrm{E}\,|W|^{\alpha}$ for some $-1 < \alpha < 0$. The important point in the context of stock markets is that the stationary ergodic martingale differences assumption is quite general and more realistic than an iid assumption. Note also that (A3a) implies that $\mathrm{E}\,\mathrm{sgn}(X-\mu)|X-\mu|^{\beta} = 0$, so if it is supported as a null hypothesis for a variety of β then the symmetry of the marginal log return distribution is also

supported.

Our final step is to study a studentised version of $g_n(\beta)$, which we define as $t_n(\beta)$. The next theorem provides this.

Theorem 6.2.3. *Assume that* $\mathrm{E}\,|X|^{2\vee\lceil 2\beta\rceil} < \infty$, *(A2) holds, and (A3a) holds with* $\beta > 1$. *Then, as* $n \to \infty$,

$$\begin{aligned} t_n(\beta) &:= T_n(\beta)/\tau_n(\beta) \qquad (6.2.8)\\ &= \frac{1}{\tau_n(\beta)\sqrt{n}} \sum_{i=1}^{n} \operatorname{sgn}\left(X_i - \overline{X}_n\right) \left|X_i - \overline{X}_n\right|^{\beta} \\ &\xrightarrow{\mathrm{D}} \mathrm{N}(0,1) \end{aligned}$$

where $\tau_n^2(\beta)$ *is obtained from* $\tau^2(\beta)$ *in* (6.2.5) *by replacing each term of the form* $E|W|^{\alpha}$ *by its consistent estimator* $\frac{1}{n}\sum_{i=1}^{n} |X_i - \overline{X}_n|^{\alpha}$.

Thus, under the assumptions of Theorem 6.2.1, the statistic $t_n(\beta)$ can be used to test the null hypothesis that skewness is zero in the usual fashion.

6.3 Simulation Results

Simulation was used to assess the rate of convergence of the distribution of $t_n(\beta)$ to a standard normal distribution. The intended empirical test is a two sided test at the 95% significance level on time series with around 1000 data points per series. Thus, for our present purposes, the most interesting property of the $t_n(\beta)$ distribution is its 2.5% quantile calculated from time series of roughly that length. In each study reported here, the quantile was determined using a sample of 50000 simulated $t_n(\beta)$ statistics with β varying from 1 to 3.

The first set of tests were performed using time series of iid normal random variables. The tests were repeated for time series of lengths 10, 100, 1000, and 10000. As can be seen in Figure 6.2, with each increase in the number of points per

series, the statistic moves closer to the theoretical value. Figure 6.3 shows the 95% confidence interval for the statistic calculated from series with 1000 data points. The theoretical value clearly lies within the confidence interval.

The second set of tests was identical to the first set, except that random variables were drawn from a Student t distribution with 4 degrees of freedom. Figure 6.4 shows the tests for varying time series lengths and Figure 6.5 shows the 95% confidence interval with the statistic calculated from time series with 1000 data points. Note that in this case the convergence results only hold for $1 \leq \beta < 2$. For $\beta \geq 2$ the fourth empirical moment does not converge to a finite limit and, being in the denominator, pulls the statistic towards zero.

For $1 < \beta < 2$, convergence of the quantiles to their theoretical limit is much slower than for the normal distribution, and there is a noticeable bias of the statistic towards lower absolute values even for sample size 1000. Nevertheless, as Figure 6.5 shows, the 95% confidence intervals contain the true quantile value for $1 < \beta < 2$.

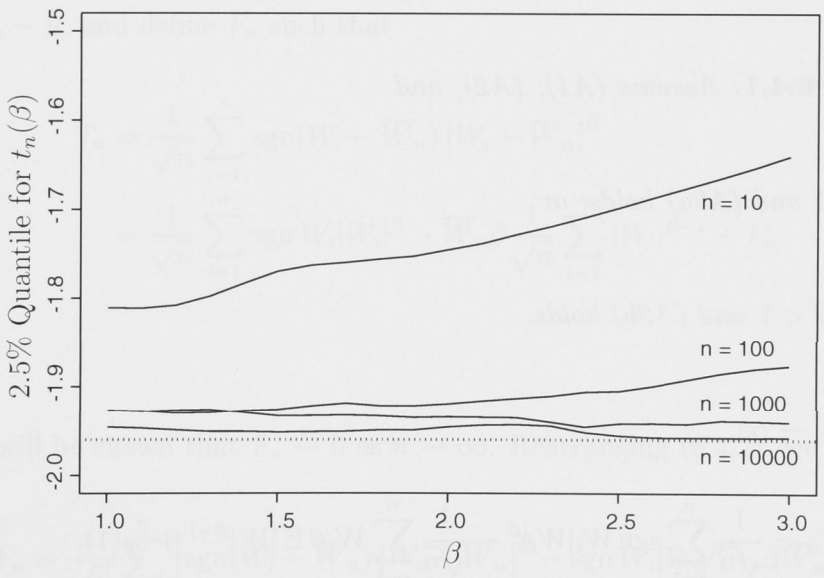

Figure 6.2: Simulated 2.5% quantiles of $t_n(\beta)$ from 50000 time series of 10, 100, 1000, and 10000 data points per series; normally distributed data. Dotted line shows theoretical limit.

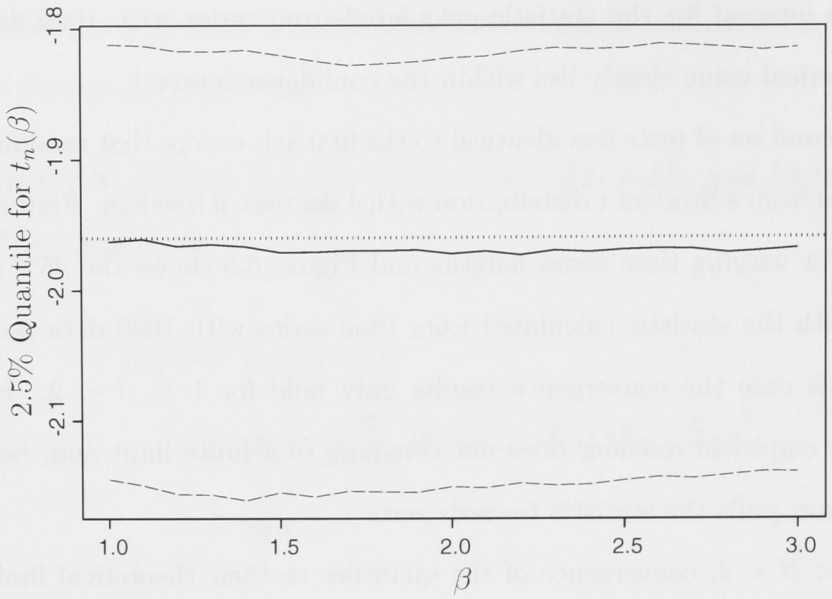

Figure 6.3: Simulated 2.5% quantiles of $t_n(\beta)$ from 50000 time series of 1000 data points per series; normally distributed data. Dashed lines show 95% confidence interval. Dotted line shows theoretical limit.

6.4 Proofs for Theorems 6.2.1, 6.2.2, and 6.2.3

Theorem 6.4.1. *Assume (A1), (A2), and*

(i) $\beta \geq 1$ *and (A3a) holds, or*

(ii) $0 < \beta < 1$ *and (A3b) holds.*

Then, as $n \to \infty$,

$$T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \operatorname{sgn} W_i |W_i|^\beta - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i \beta \, \mathrm{E}\, |W|^{\beta-1} + o_p(1). \tag{6.4.1}$$

Proof. Assume (A1) and (A2).

(i) (Martingale Case) Assume in addition that $\beta \geq 1$ and (A3a) holds. Let $\overline{W}_n =$

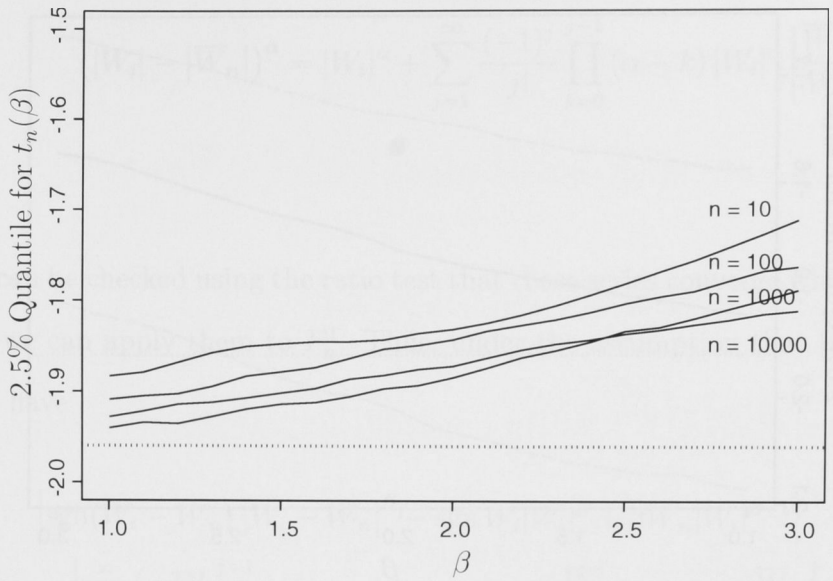

Figure 6.4: Simulated 2.5% quantiles of $t_n(\beta)$ from 50000 time series of 10, 100, 1000, and 10000 data points per series; t_4 distributed data. Dotted line shows theoretical limit.

$\overline{X}_n - \mu$, and define F_n such that

$$\begin{aligned} T_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \operatorname{sgn}(W_i - \overline{W}_n) \left| W_i - \overline{W}_n \right|^{\beta} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \operatorname{sgn} W_i |W_i|^{\beta} - \overline{W}_n \beta \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |W_i|^{\beta-1} + F_n. \end{aligned} \tag{6.4.2}$$

It will be shown that $F_n \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$. Rearranging (6.4.2), we see that

$$F_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\operatorname{sgn}(W_i - \overline{W}_n) \left| W_i - \overline{W}_n \right|^{\beta} - \operatorname{sgn} W_i |W_i|^{\beta} + \beta \overline{W}_n |W_i|^{\beta-1} \right].$$

The case $\beta = 1$ is trivial, so from now on we keep $\beta > 1$. We decompose F_n

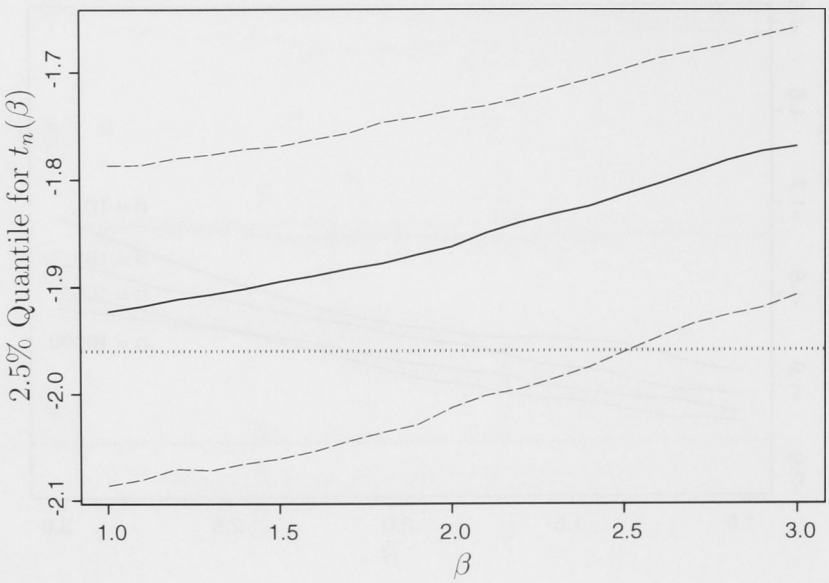

Figure 6.5: Simulated 2.5% quantiles of $t_n(\beta)$ from 50000 time series of 1000 data points per series; t_4 distributed data. Dashed lines show 95% confidence interval. Dotted line shows theoretical limit.

into two partial sums, F_n^1 and F_n^2:

$$\begin{aligned} F_n^1 &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{1}\left(|W_i| > \left|\overline{W}_n\right|\right) \\ &\qquad \times \left[\operatorname{sgn}(W_i - \overline{W}_n)\left|W_i - \overline{W}_n\right|^{\beta} - \operatorname{sgn} W_i |W_i|^{\beta} + \beta \overline{W}_n |W_i|^{\beta-1}\right]; \\ F_n^2 &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{1}\left(|W_i| \leq \left|\overline{W}_n\right|\right) \\ &\qquad \times \left[\operatorname{sgn}(W_i - \overline{W}_n)\left|W_i - \overline{W}_n\right|^{\beta} - \operatorname{sgn} W_i |W_i|^{\beta} + \beta \overline{W}_n |W_i|^{\beta-1}\right], \end{aligned}$$

which will be shown to be negligible as $n \to \infty$.

The following Taylor expansions in terms of $\overline{W}_n$ will be useful:

$$\begin{aligned} &\operatorname{sgn}(W_i - \overline{W}_n)\left|W_i - \overline{W}_n\right|^{\beta} \\ &= \operatorname{sgn} W_i |W_i|^{\beta} + \sum_{j=1}^{\infty} \frac{(-1)^j}{j!} \prod_{k=0}^{j-1} (\beta - k) \operatorname{sgn}^{j-1} W_i |W_i|^{\beta} \left(\frac{\overline{W}_n}{|W_i|}\right)^j, \end{aligned}$$

and, for $\alpha > 0$,

$$\left(|W_i| - \left|\overline{W}_n\right|\right)^\alpha = |W_i|^\alpha + \sum_{j=1}^{\infty} \frac{(-1)^j}{j!} \prod_{k=0}^{j-1} (\alpha - k) \, |W_i|^\alpha \left|\frac{\overline{W}_n}{W_i}\right|^j .$$

It can be checked using the ratio test that these series converge when $\left|\frac{\overline{W}_n}{W_i}\right| < 1$, so we can apply them to F_n^1. Thus, under the assumption that $|W_i| > \left|\overline{W}_n\right|$, we have

$$\begin{aligned}
&\left|\operatorname{sgn}(W_i - \overline{W}_n) \left|W_i - \overline{W}_n\right|^\beta - \operatorname{sgn} W_i |W_i|^\beta + \beta \overline{W}_n |W_i|^{\beta-1}\right| \\
&= \left|\sum_{j=1}^{\infty} \frac{(-1)^j}{j!} \prod_{k=0}^{j-1} (\beta - k) \operatorname{sgn}^{j-1} W_i |W_i|^\beta \frac{\overline{W}_n^j}{|W_i|^j} + \beta |W_i|^\beta \frac{\overline{W}_n}{|W_i|}\right| \\
&= \left|\sum_{j=2}^{\infty} \frac{(-1)^j}{j!} \prod_{k=0}^{j-1} (\beta - k) \operatorname{sgn}^{j-1} W_i |W_i|^\beta \frac{\overline{W}_n^j}{|W_i|^j}\right| \\
&\le \sum_{j=2}^{\lfloor\beta\rfloor} \frac{1}{j!} \prod_{k=0}^{j-1} (\beta - k) |W_i|^\beta \left|\frac{\overline{W}_n}{W_i}\right|^j + \sum_{j=\lfloor\beta\rfloor+1}^{\infty} \frac{1}{j!} \prod_{k=0}^{j-1} |\beta - k| |W_i|^\beta \left|\frac{\overline{W}_n}{W_i}\right|^j .
\end{aligned}$$

The second sum on the right hand side does not exceed

$$\begin{aligned}
&\left|\overline{W}_n\right|^{\lfloor\beta\rfloor} \prod_{m=0}^{\lfloor\beta\rfloor-1} (\beta - m) \sum_{j=1}^{\infty} \frac{1}{(j + \lfloor\beta\rfloor)!} \prod_{k=0}^{j-1} |\beta - \lfloor\beta\rfloor - k| \, |W_i|^{\beta - \lfloor\beta\rfloor} \left|\frac{\overline{W}_n}{W_i}\right|^j \\
&\le -\left|\overline{W}_n\right|^{\lfloor\beta\rfloor} \prod_{m=0}^{\lfloor\beta\rfloor-1} (\beta - m) \sum_{j=1}^{\infty} \frac{(-1)^j}{j!} \prod_{k=0}^{j-1} (\beta - \lfloor\beta\rfloor - k) \, |W_i|^{\beta - \lfloor\beta\rfloor} \left|\frac{\overline{W}_n}{W_i}\right|^j \\
&\le -\left|\overline{W}_n\right|^{\lfloor\beta\rfloor} \prod_{m=0}^{\lfloor\beta\rfloor-1} (\beta - m) \left[\left(|W_i| - \left|\overline{W}_n\right|\right)^{\beta - \lfloor\beta\rfloor} - |W_i|^{\beta - \lfloor\beta\rfloor}\right] \\
&\le \beta^\beta |W_i|^{\beta - \lfloor\beta\rfloor - 1} \left|\overline{W}_n\right|^{\lfloor\beta\rfloor + 1} .
\end{aligned}$$

The final inequality follows because

$$\prod_{m=0}^{\lfloor\beta\rfloor-1} (\beta - m) \le \beta^\beta ,$$

and $0 \leq \beta - \lfloor\beta\rfloor < 1$, so

$$|W_i|^{\beta-\lfloor\beta\rfloor} - \left(|W_i| - \left|\overline{W}_n\right|\right)^{\beta-\lfloor\beta\rfloor} = |W_i|^{\beta-\lfloor\beta\rfloor}\left[1 - \left(1 - \left|\frac{\overline{W}_n}{W_i}\right|\right)^{\beta-\lfloor\beta\rfloor}\right]$$
$$\leq |W_i|^{\beta-\lfloor\beta\rfloor-1}\left|\overline{W}_n\right|.$$

Incorporating the above result into F_n^1 gives

$$\begin{aligned}
\left|F_n^1\right| &\leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\sum_{j=2}^{\lfloor\beta\rfloor}\frac{1}{j!}\prod_{k=0}^{j-1}(\beta-k)|W_i|^{\beta}\left|\frac{\overline{W}_n}{W_i}\right|^{j} + \beta^{\beta}|W_i|^{\beta-\lfloor\beta\rfloor-1}\left|\overline{W}_n\right|^{\lfloor\beta\rfloor+1}\right] \\
&\leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=2}^{\lfloor\beta\rfloor}\frac{1}{j!}\prod_{k=0}^{j-1}(\beta-k)|W_i|^{\beta}\left|\frac{\overline{W}_n}{W_i}\right|^{j} + \frac{\beta^{\beta}}{\sqrt{n}}\sum_{i=1}^{n}|W_i|^{\beta-\lfloor\beta\rfloor-1}\left|\overline{W}_n\right|^{\lfloor\beta\rfloor+1} \\
&\leq \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_i\right|\sum_{j=2}^{\lfloor\beta\rfloor}\frac{1}{j!}\prod_{k=0}^{j-1}(\beta-k)\left|\overline{W}_n\right|^{j-1}\frac{1}{n}\sum_{i=1}^{n}|W_i|^{\beta-j} \\
&\quad + \left|\frac{\beta^{\beta}}{\sqrt{n}}\sum_{i=1}^{n}W_i\right|\left|\overline{W}_n\right|^{\lfloor\beta\rfloor}\frac{1}{n}\sum_{i=1}^{n}|W_i|^{\beta-\lfloor\beta\rfloor-1}.
\end{aligned}$$

This can now be evaluated as $n \to \infty$. Under assumption (A1) we have $\mathrm{E}\,|W|^2 < \infty$. Also, $\mathrm{E}(W_i|\mathcal{F}_{i-1}) = 0$ because $\{W_i\}$ are martingale differences. The Central Limit Theorem (CLT) for stationary ergodic martingale differences (Heyde and Hall 1980, p. 10) then gives

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(W_i - \mathrm{E}(W_i|\mathcal{F}_{i-1})\right) \xrightarrow{\mathrm{D}} \mathrm{N}\left(0, \mathrm{E}\,W^2\right).$$

Therefore, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_i \xrightarrow{\mathrm{D}} \mathrm{N}(0, \mathrm{E}\,W^2)$. Also, $\mathrm{E}\,W = 0$, so by the Ergodic Theorem, $\overline{W}_n \xrightarrow{\text{a.s.}} 0$.

Now we need the following:

Lemma 6.4.1. *Under (A1) and (A2),* $\mathrm{E}\,|W + x|^b < \infty$, *where* $-1 < b < 0$ *and* $|x| < \frac{1}{2}$.

Proof. By assumption (A2), we have

$$c := \sup_{w \in \mathbb{R}} f_W(w) = \sup_{x \in \mathbb{R}} f_X(x) < \infty,$$

where $f_W(w)$ is the probability density function for W. Thus for $|x| < \frac{1}{2}$ and $-1 < b < 0$,

$$\begin{aligned}
\mathrm{E}\,|W+x|^b &= \int_{|w|<1} |w+x|^b f_W(w)\mathrm{d}w + \int_{|w|\geq 1} |w+x|^b f_W(w)\mathrm{d}w \\
&\leq c \int_{|w|<1} |w+x|^b \mathrm{d}w + 2^{-b} \int_{|w|\geq 1} f_W(w)\mathrm{d}w \\
&\leq \frac{c}{b+1}\left[(1-x)^{b+1} + (1+x)^{b+1}\right] + 2^{-b} \\
&\leq \frac{2^{b+2}c}{b+1} + 2^{-b}.
\end{aligned}$$

□

Lemma 6.4.1 and (A1) together imply that $\mathrm{E}\,|X|^\gamma < \infty$ for $-1 < \gamma \leq 2\beta$, so again by the Ergodic Theorem,

$$\frac{1}{n}\sum_{i=1}^{n} |W_i|^{\beta - \lfloor\beta\rfloor - 1} \xrightarrow{\text{a.s.}} \mathrm{E}\,|W|^{\beta-\lfloor\beta\rfloor-1} \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} W_i^{\beta-j} \xrightarrow{\text{a.s.}} \mathrm{E}\,W^{\beta-j}$$

for $j \in \{2, \ldots, \lfloor\beta\rfloor\}$. Thus we have shown that $F_n^1 \xrightarrow{\text{P}} 0$.

F_n^2 requires less effort. Using the CLT for stationary ergodic martingale differences and the Ergodic Theorem, as before, we have

$$\begin{aligned}
\left|F_n^2\right| &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbf{1}\left(|W_i| \leq \left|\overline{W}_n\right|\right)\Big[\operatorname{sgn}(W_i - \overline{W}_n)\left|W_i - \overline{W}_n\right|^\beta \\
&\qquad - \operatorname{sgn} W_i |W_i|^\beta + \beta \overline{W}_n |W_i|^{\beta-1}\Big] \\
&\leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbf{1}\left(|W_i| \leq \left|\overline{W}_n\right|\right)\left[\left|W_i - \overline{W}_n\right|^\beta + |W_i|^\beta + \beta\left|\overline{W}_n\right| |W_i|^{\beta-1}\right]
\end{aligned}$$

$$\leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[2^{\beta} \left| \overline{W}_n \right|^{\beta} + \left| \overline{W}_n \right|^{\beta} + \beta \left| \overline{W}_n \right|^{\beta} \right]$$

$$\leq \left(2^{\beta} + \beta + 1 \right) \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i \right| \left| \overline{W}_n \right|^{\beta - 1} \xrightarrow{\text{P}} 0.$$

Consolidating the above results we deduce that $|F_n| \leq |F_n^1| + |F_n^2| \xrightarrow{\text{P}} 0$. To complete the proof of Part (i) of the Theorem, note that the Ergodic Theorem gives that $\frac{1}{n} \sum_{i=1}^{n} |W_i|^{\beta-1} \xrightarrow{\text{a.s.}} \mathrm{E}\,|W|^{\beta-1}$ when $\beta \geq 1$. Hence, using Equation (6.4.2), we obtain Equation (6.4.1) from

$$\begin{aligned} T_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \operatorname{sgn} W_i |W_i|^{\beta} - \overline{W}_n \frac{\beta}{\sqrt{n}} \sum_{i=1}^{n} |W_i|^{\beta-1} + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \operatorname{sgn} W_i |W_i|^{\beta} - \left(\frac{\beta}{\sqrt{n}} \sum_{i=1}^{n} W_i \right) \left(\frac{1}{n} \sum_{i=1}^{n} |W_i|^{\beta-1} \right) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \operatorname{sgn} W_i |W_i|^{\beta} - \mathrm{E}\,|W|^{\beta-1} \frac{\beta}{\sqrt{n}} \sum_{i=1}^{n} W_i + o_p(1). \end{aligned}$$

(ii) (iid Case) Assume instead that $0 < \beta < 1$ and (A3b) holds. Then again we will show that Equation (6.4.1) holds as $n \to \infty$. To see this, define G_n so that

$$\sum_{i=1}^{n} \operatorname{sgn}(X_i - \overline{X}_n) |X_i - \overline{X}_n|^{\beta} = \sum_{i=1}^{n} \operatorname{sgn}(W_i) |W_i|^{\beta} + \beta G_n,$$

where, as before, $W_i = X_i - \mu$. Thus, since $\overline{W}_n = \overline{X}_n - \mu$,

$$\begin{aligned} G_n &= \sum_{i=1}^{n} \int_0^{-\overline{W}_n} |W_i + x|^{\beta-1} \, \mathrm{d}x \\ &= n \int_0^{-\overline{W}_n} \mathrm{E}\,|W + x|^{\beta-1} \, \mathrm{d}x + \int_0^{-\overline{W}_n} S_n(x) \mathrm{d}x, \qquad (6.4.3) \end{aligned}$$

$$\text{where} \quad S_n(x) = \sum_{i=1}^{n} \left(|W_i + x|^{\beta-1} - \mathrm{E}\,|W + x|^{\beta-1} \right).$$

It will be shown that the second integral in Equation (6.4.3) converges in probability to zero. This will be proven separately for the cases $\frac{1}{2} < \beta < 1$ and

$0 < \beta \leq \frac{1}{2}$.

(iia) (Case $\frac{1}{2} < \beta < 1$) Set $A = \{0 \wedge -\overline{W}_n \leq x \leq 0 \vee -\overline{W}_n\}$, $\varepsilon > 0$, and $0 < \delta < \frac{1}{2}$. Note that

$$\begin{aligned} P_1(n) &:= \mathrm{P}\left(\frac{1}{\sqrt{n}}\left|\int_0^{-\overline{W}_n} S_n(x)\mathrm{d}x\right| > \varepsilon\right) \\ &\leq \mathrm{P}\left(\frac{1}{\sqrt{n}}\left|\int_{|x|\leq\delta} \mathbf{1}_A\, S_n(x)\mathrm{d}x\right| > \varepsilon, \left|\overline{W}_n\right| \leq \delta\right) + \mathrm{P}\left(\left|\overline{W}_n\right| > \delta\right) \\ &\leq \mathrm{P}\left(\left|\int_{|x|\leq\delta} \mathbf{1}_A\, S_n(x)\mathrm{d}x\right| > \varepsilon\sqrt{n}\right) + \frac{1}{\delta^2 n}\,\mathrm{E}\,W^2, \end{aligned}$$

by Chebyshev's Inequality. Now, by the Cauchy-Schwarz Inequality,

$$\begin{aligned} P_2(n) &:= \mathrm{P}\left(\left|\int_{|x|\leq\delta} \mathbf{1}_A\, S_n(x)\mathrm{d}x\right| > \varepsilon\sqrt{n}\right) \\ &\leq \frac{1}{\varepsilon^2 n}\,\mathrm{E}\left(\left[\int_{|x|\leq\delta} \mathbf{1}_A\, S_n(x)\mathrm{d}x\right]^2\right) \\ &\leq \frac{1}{\varepsilon^2 n}\,\mathrm{E}\left(\int_{|x|\leq\delta} S_n^2(x)\mathrm{d}x \int_{|x|\leq\delta} \mathbf{1}_A\,\mathrm{d}x\right) \\ &\leq \frac{2\delta}{\varepsilon^2 n}\int_{|x|\leq\delta} \mathrm{E}\left(S_n^2(x)\right)\mathrm{d}x \\ &= \frac{2\delta}{\varepsilon^2}\int_{|x|\leq\delta} \mathrm{var}\,|W+x|^{\beta-1}\,\mathrm{d}x \\ &\leq \frac{2\delta}{\varepsilon^2}\int_{|x|\leq\delta} \mathrm{E}\,|W+x|^{2(\beta-1)}\,\mathrm{d}x. \end{aligned} \tag{6.4.4}$$

Note that the iid assumption gives that $\mathrm{E}\,(S_n^2(x)) = \mathrm{var}\,|W+x|^{\beta-1}$.

Next, recall that $\delta < \frac{1}{2}$ and define

$$K_\delta = \sup_{|x|\leq\delta} \mathrm{E}\,|W+x|^{2(\beta-1)},$$

which is finite by Lemma 6.4.1 since $\frac{1}{2} < \beta < 1$.

Choosing $\delta = an^{-\frac{1}{2}}$, where $a > 0$, yields

$$P_1(n) \leq \frac{4a^2}{\varepsilon^2 n} K_{an^{-\frac{1}{2}}} + \frac{1}{a^2} \mathrm{E}\, W^2.$$

So $\limsup_{n\to\infty} P_1(n) \leq \frac{1}{a^2} \mathrm{E}\, W^2$. Allowing $a \to \infty$ then shows that $P_1(n) \to 0$ as $n \to \infty$.

(iib) (Case $0 < \beta \leq \frac{1}{2}$) This case needs to be addressed separately as Equation (6.4.4) ceases to be useful when $\mathrm{E}\,|W + x|^{2(\beta-1)}$ is unbounded. To find an alternative argument, take $\frac{1}{2} < q \leq 1$ and choose $q < \frac{1}{2(1-\beta)}$, as is possible since $0 < \beta < \frac{1}{2}$, implying $\frac{1}{2} < \frac{1}{2(1-\beta)} < 1$. Note then that

$$\begin{aligned} P_2(n) &= \mathrm{P}\left(\left|\int_{|x|\leq\delta} \mathbf{1}_A\, S_n(x)\mathrm{d}x\right| > \varepsilon\sqrt{n}\right) \\ &= \mathrm{P}\left(\left|\int_{|x|\leq\delta} \mathbf{1}_A\, S_n(x)\mathrm{d}x\right|^{2q} > \varepsilon^{2q} n^q\right) \\ &\leq \frac{1}{\varepsilon^{2q} n^q} \mathrm{E}\left|\int_{|x|\leq\delta} \mathbf{1}_A\, S_n(x)\mathrm{d}x\right|^{2q}, \end{aligned}$$

by the Markov Inequality. Now, noting that $\frac{1}{2} < q \leq 1$ and using Hölder's Inequality,

$$\begin{aligned} \left|\int_{|x|\leq\delta} \mathbf{1}_A\, S_n(x)\mathrm{d}x\right|^{2q} &\leq \left[\int_{|x|\leq\delta} |\mathbf{1}_A\, S_n(x)|\,\mathrm{d}x\right]^{2q} \\ &\leq \left[\int_{|x|\leq\delta} |\mathbf{1}_A|^{\frac{2q}{2q-1}}\,\mathrm{d}x\right]^{2q-1} \int_{|x|\leq\delta} |S_n(x)|^{2q}\,\mathrm{d}x \\ &\leq (2\delta)^{2q-1} \int_{|x|\leq\delta} |S_n(x)|^{2q}\,\mathrm{d}x. \end{aligned}$$

Thus,

$$P_2(n) \leq \frac{(2\delta)^{2q-1}}{\varepsilon^{2q} n^q} \mathrm{E} \int_{|x|\leq\delta} |S_n(x)|^{2q}\,\mathrm{d}x.$$

Now we use the iid assumption again to apply an inequality of von Bahr

and Esseen (Petrov 1995, p. 82), which gives, for $|x| \leq \delta$,

$$\mathrm{E}\,|S_n(x)|^{2q} \leq 2n\,\mathrm{E}\left||W+x|^{\beta-1} - \mathrm{E}\,|W+x|^{\beta-1}\right|^{2q} \leq 2nC_\delta,$$

where

$$C_\delta = \sup_{|x|\leq\delta} \mathrm{E}\left||W+x|^{\beta-1} - \mathrm{E}\,|W+x|^{\beta-1}\right|^{2q}.$$

C_δ is finite because

$$\begin{aligned}\mathrm{E}\left||W+x|^{\beta-1} - \mathrm{E}\,|W+x|^{\beta-1}\right|^{2q} &\leq \mathrm{E}\,|W+x|^{2q(\beta-1)} + \left[\mathrm{E}\,|W+x|^{\beta-1}\right]^{2q} \\ &\leq 2\,\mathrm{E}\,|W+x|^{2q(\beta-1)}, \qquad (6.4.5)\end{aligned}$$

by Jensen's Inequality. We chose $\delta < \frac{1}{2}$ and $\frac{1}{2} < q \leq 1$, so that $0 < q < \frac{1}{2(1-\beta)}$, thus $-1 < -2q(1-\beta) < 0$, and thus by Lemma 6.4.1, the final expectation in Equation (6.4.5) is finite.

Repeating the substitution $\delta = an^{-\frac{1}{2}}$ gives

$$P_2(n) \leq \frac{(2\delta)^{2q-1}}{\varepsilon^{2q}n^q} 4n\delta C_\delta = \frac{2a^{2q}n^{1-2q}C_\delta}{\varepsilon^{2q}}$$

and $P_1(n)$ becomes

$$P_1(n) \leq \frac{2a^{2q}n^{1-2q}C_{an^{-\frac{1}{2}}}}{\varepsilon^{2q}} + \frac{1}{a^2}\,\mathrm{E}\,W^2.$$

Once more $\limsup_{n\to\infty} P_1(n) \leq \frac{1}{a^2}\,\mathrm{E}\,W^2$ and allowing $a \to \infty$ then shows that $P_1(n) \to 0$ as $n \to \infty$.

Thus it has been shown that in the iid Case, for $0 < \beta < 1$,

$$\frac{1}{\sqrt{n}} \int_0^{-\overline{W}_n} S_n(x)\mathrm{d}x \xrightarrow{\mathrm{P}} 0.$$

The next step is to show that

$$\frac{1}{-\overline{W}_n}\int_0^{-\overline{W}_n} \mathrm{E}\,|W+x|^{\beta-1}\,\mathrm{d}x \xrightarrow{\text{a.s.}} \mathrm{E}\,|W|^{\beta-1}. \tag{6.4.6}$$

By the Dominated Convergence Theorem,

$$\lim_{t\to 0}\frac{1}{t}\int_0^t \mathrm{E}\,|W+x|^{\beta-1}\,\mathrm{d}x = \lim_{t\to 0}\int_0^1 \mathrm{E}\,|W+tx|^{\beta-1}\,\mathrm{d}x = \mathrm{E}\,|W|^{\beta-1}.$$

Note that $\sup_{0\le x\le 1}\mathrm{E}\,|W+tx| \le C$ for a finite constant C, once $t \le \frac{1}{2}$, by Lemma 6.4.1. Since $\overline{W}_n \xrightarrow{\text{a.s.}} 0$ as $n\to\infty$, we have $\overline{W}_n(\omega)\to 0$ for each $\omega\in\Omega'$, where $\Omega'\subseteq\Omega$ and $\mathrm{P}(\Omega')=1$. Thus Equation (6.4.6) holds.

So it has been shown that

$$\begin{aligned}\frac{G_n}{\sqrt{n}} &= -\sqrt{n}\overline{W}_n\frac{1}{-\overline{W}_n}\int_0^{-\overline{W}_n}\mathrm{E}\,|W+x|^{\beta-1}\,\mathrm{d}x + \frac{1}{\sqrt{n}}\int_0^{-\overline{W}_n} S_n(x)\mathrm{d}x\\ &= -\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i\,\mathrm{E}\,|W|^{\beta-1}+o_p(1).\end{aligned}$$

Since

$$T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathrm{sgn}(W_i)|W_i|^\beta - \frac{\beta}{\sqrt{n}}G_n,$$

we see that Equation (6.4.1) holds again. □

Proof of Theorem 6.2.1. To show that Equation 6.2.4 holds, the Central Limit Theorem for stationary ergodic martingale differences is applied to Equation 6.4.1. Note that it applies equally well in the iid case. By assumption (A1), for any real constant c, $\mathrm{E}\left(\mathrm{sgn}\,W|W|^\beta + cW\right)^2 < \infty$, so

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\mathrm{sgn}\,W_i|W_i|^\beta + cW_i - \mathrm{E}\left(\mathrm{sgn}\,W_i|W_i|^\beta + cW_i\middle|\,\mathcal{F}_{i-1}\right)\right)\\ &\xrightarrow{\text{D}} \mathrm{N}\left(0, \mathrm{E}\left(\mathrm{sgn}\,W|W|^\beta + cW\right)^2\right).\end{aligned}$$

By (A3a) or by (A3b), for $i \in \{1, \ldots, n\}$

$$\mathrm{E}\left(\operatorname{sgn} W_i |W_i|^\beta + cW_i \middle| \mathcal{F}_{i-1}\right) = \mathrm{E}\left(\operatorname{sgn} W_i |W_i|^\beta \middle| \mathcal{F}_{i-1}\right) + c\,\mathrm{E}\left(W_i \middle| \mathcal{F}_{i-1}\right) = 0.$$

Also,

$$\begin{aligned}\mathrm{E}\left(\operatorname{sgn} W|W|^\beta + cW\right)^2 &= \mathrm{E}\,|W|^{2\beta} + 2c\,\mathrm{E}\left(\operatorname{sgn}(W)W|W|^\beta\right) + c^2\,\mathrm{E}\,W^2 \\ &= \mathrm{E}\,|W|^{2\beta} + 2c\,\mathrm{E}\,|W|^{\beta+1} + c^2\,\mathrm{E}\,W^2.\end{aligned}$$

Thus

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\operatorname{sgn} W_i |W_i|^\beta + cW_i\right) \xrightarrow{\mathrm{D}} \mathrm{N}\left(0, \mathrm{E}\,|W|^{2\beta} + 2c\,\mathrm{E}\,|W|^{\beta+1} + c^2\,\mathrm{E}\,W^2\right).$$

Substituting $c = -\beta\,\mathrm{E}\,|W|^{\beta-1}$ and appealing to the Cramer-Wold device gives $T_n \xrightarrow{\mathrm{D}} \mathrm{N}\left(0, \tau^2(\beta)\right)$ as $n \to \infty$ where $\tau^2(\beta)$ is given in (6.2.5). □

Proof of Theorem 6.2.2. Again define $W_i = X_i - \mu$ for $i \in \{1, \ldots, n\}$ and note that $X_i - \overline{X}_n = W_i - \overline{W}_n$. The proof is in two parts for different values of α.

(i) Constrain α such that $0 \le \alpha \le 1$ and note that for real a and b,

$$\left||a+b|^\alpha - |a|^\alpha\right| = \left|\int_{|a|}^{|a+b|} \alpha x^{\alpha-1} \mathrm{d}x\right| \le \left|\int_{|0|}^{|b|} \alpha x^{\alpha-1} \mathrm{d}x\right| \le |b|^\alpha.$$

So, as $n \to \infty$,

$$\left|\frac{1}{n}\sum_{i=1}^{n} \left|W_i - \overline{W}_n\right|^\alpha - |W_i|^\alpha\right| \le \frac{1}{n}\sum_{i=1}^{n} \left|\left|W_i - \overline{W}_n\right|^\alpha - |W_i|^\alpha\right| \le \left|\overline{W}_n\right|^\alpha \xrightarrow{\text{a.s.}} 0.$$

(ii) Next set $1 \le \alpha \le 2 \vee \lceil 2\beta \rceil$. In this case, by the Minkowski inequality,

$$\left(\sum_{i=1}^{n} \left|W_i - \overline{W}_n\right|^\alpha\right)^{\frac{1}{\alpha}} \le \left(\sum_{i=1}^{n} |W_i|^\alpha\right)^{\frac{1}{\alpha}} + \left(\sum_{i=1}^{n} \left|\overline{W}_n\right|^\alpha\right)^{\frac{1}{\alpha}}$$

so

$$\left(\frac{1}{n}\sum_{i=1}^{n}\left|W_i-\overline{W}_n\right|^{\alpha}\right)^{\frac{1}{\alpha}}-\left(\frac{1}{n}\sum_{i=1}^{n}\left|W_i\right|^{\alpha}\right)^{\frac{1}{\alpha}}\leq\left|\overline{W}_n\right|.$$

Similarly,

$$\left(\frac{1}{n}\sum_{i=1}^{n}\left|W_i\right|^{\alpha}\right)^{\frac{1}{\alpha}}-\left(\frac{1}{n}\sum_{i=1}^{n}\left|W_i-\overline{W}_n\right|^{\alpha}\right)^{\frac{1}{\alpha}}\leq\left|\overline{W}_n\right|,$$

and thus

$$\left|\left(\frac{1}{n}\sum_{i=1}^{n}\left|W_i-\overline{W}_n\right|^{\alpha}\right)^{\frac{1}{\alpha}}-\left(\frac{1}{n}\sum_{i=1}^{n}\left|W_i\right|^{\alpha}\right)^{\frac{1}{\alpha}}\right|\leq\left|\overline{W}_n\right|\xrightarrow{\text{a.s.}}0,\quad n\to\infty.$$

So for all $0\leq\alpha\leq 2\vee\lceil 2\beta\rceil$,

$$\frac{1}{n}\sum_{i=1}^{n}\left|X_i-\overline{X}_n\right|^{\alpha}\xrightarrow{\text{a.s.}}\frac{1}{n}\sum_{i=1}^{n}\left|X_i-\mu\right|^{\alpha}\xrightarrow{\text{a.s.}}\mathrm{E}\left|X-\mu\right|^{\alpha}$$

as $n\to\infty$. □

Proof of Theorem 6.2.3. By Theorem 6.2.1 it suffices to show that $\tau_n(\beta)\xrightarrow{\text{P}}\tau(\beta)$ as $n\to\infty$. This follows from Theorem 6.2.2, as its assumptions encompass both (A3a) and (A3b). □

6.5 A Two Stage Test for Skewness

Theorem 6.2.3 allows us to construct a test statistic for the null hypothesis that the marginal distribution of our sample $\{X_i\}_{i\in\{1,\ldots,n\}}$ is symmetric, given that our drift corrected samples are stationary ergodic martingale differences with finite variance. There is an alternative route to this result. Consider the following theorems.

Theorem 6.5.1. *Assume we are given a sample of random variables* $\{X_i\}_{i\in\{1,\dots,n\}}$ *defined on a filtered probability space* $(\Omega, \mathcal{F}_n, \mathrm{P})$. *Take* $\beta > 0$ *and assume that*

(B1) $\mathrm{E}\, X^{2\beta} < \infty$,

(B2) X *has probability density function* $f_X(x)$ *satisfying* $\sup_{x\in\mathbb{R}} f_X(x) < \infty$,

(B3) $\mu = \mathrm{E}\, X = 0$, *and that*

(B4) $\{X_i\}_{i\in\{1,\dots,n\}}$ *are stationary ergodic martingale differences, having*

(B5) $\mathrm{E}\left(\operatorname{sgn}(X_i)|X_i|^\beta \middle| \mathcal{F}_{i-1}\right) = 0$ *for all* $i \in \{1,\dots,n\}$.

Then the following holds as $n \to \infty$*:*

$$T_n^A(\beta) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \operatorname{sgn}(X_i)\, |X_i|^\beta \xrightarrow{\mathrm{D}} \mathrm{N}\left(0, \mathrm{E}\,|X|^{2\beta}\right). \tag{6.5.1}$$

Proof. The Central Limit Theorem for stationary ergodic martingale differences (Heyde and Hall 1980, p. 10) and assumptions (B1) and (B4) give that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\operatorname{sgn} X_i\, |X_i|^\beta - \mathrm{E}\left(\operatorname{sgn} X_i\, |X_i|^\beta \middle| \mathcal{F}_{i-1}\right)\right] \xrightarrow{d} \mathrm{N}\left(0, \mathrm{E}\,|X|^{2\beta}\right),$$

which in conjunction with assumption (B5) is the required result. □

Theorem 6.5.2. *Under the assumptions of Theorem 6.5.1, as* $n \to \infty$,

$$t_n^A(\beta) := \frac{T_n^A(\beta)}{(\mathrm{E}\,|X|^{2\beta})^{\frac{1}{2}}} = \frac{\sum_{i=1}^n \operatorname{sgn}(X_i)\, |X_i|^\beta}{(\sum_{i=1}^n |X_i|^{2\beta})^{\frac{1}{2}}} \xrightarrow{\mathrm{D}} \mathrm{N}(0,1). \tag{6.5.2}$$

Proof. The result follows from Theorem 6.5.1 and the Ergodic Theorem, which provides that $\frac{1}{n}\sum_{i=1}^n |X_i|^{2\beta} \xrightarrow{\text{a.s.}} \mathrm{E}\,|X|^{2\beta}$ as $n \to \infty$. □

Theorem 6.5.2 is the analogue of Theorem 6.2.3 under different assumptions. For Theorem 6.2.3, β is restricted so that $\mathrm{E}\,|X|^{\lceil 2\beta \rceil} < \infty$ and $\beta > 1$ but μ can take any

value. For Theorem 6.5.2, the restrictions are that $\mathrm{E}\,|X|^{2\vee 2\beta} < \infty$ and $\beta > 0$ but $\mu = 0$.

The assumption that a stock's log returns have no drift may seem to be unrealistic, but it is not uncommonly observed in financial markets, especially in series over short intervals of time. Thus this test is a two stage statistical test. The first hypothesis is $H_0^1 : \mu = 0$, $H_A^1 : \mu \neq 0$, which can be tested with a standard t-test. Note that the statistic

$$s_n = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i}{\left(\frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n \right)^2 \right)^{\frac{1}{2}}} \tag{6.5.3}$$

follows a t-distribution with $n-1$ degrees of freedom under the assumption of iid samples with finite variance, or it can be shown that $s_n \xrightarrow{\mathrm{D}} \mathrm{N}(0,1)$ as $n \to \infty$ using Theorem 6.2.2 and the Central Limit Theorem for stationary ergodic martingale differences (Heyde and Hall 1980, p. 10) if assumptions (B1–5) hold for $\beta = 1$. If a particular sample fails to reject H_0^1, then we can apply a second test $H_0^2 : g(\beta) = 0$, $H_A^2 : g(\beta) \neq 0$ using the results of Theorem 6.5.2.

The single stage test that Theorem 6.2.3 permits is simpler and deals with μ explicitly. However, the procedure presented in this section has the following drawback. If a t-test fails to reject H_0^1 but then H_0^2 is rejected, we are not able to quantify how much a small non-zero μ may have effected the second stage of the test. Further, if H_0^1 is rejected, we can say nothing about the symmetry of the sample's distribution. Ultimately the primary interest of this study became the tests associated with Theorem 6.2.3, and these are what we will use for the empirical investigations in Chapter 7.

Chapter 7

Skewness in the Australian Stock Exchange

7.1 Skewness in Finance

If returns are asymmetrically distributed, the implications for asset allocation and derivatives pricing are immediate. Recent studies where asymmetry statistics were used in asset allocation decisions include those of Patton (2004), Harvey, Liechty, Liechty, and Mueller (2004), and Korkie, Sivakumar, and Turtle (2006). While the techniques used in these papers varied considerably, all attempted to select better portfolio weights using information about the individual asymmetry and joint asymmetry of the distributions of stock returns, where joint asymmetry refers to some form of correlation amongst extreme returns between stocks. Korkie et al. warn against the use of an asymmetry statistic based on the conditional centralised third moment in the context of a particular autoregressive conditional heteroskedasticity model, as a *variance spillover* effect caused by varying conditional variance can confound the measurement of skewness in this way. More information on asymmetry and asset allocation can be found in the references given in Korkie et al. (2006).

Two leading alternatives to GBM for option pricing, Variance Gamma processes

(Madan, Carr, and Chang 1998) and stochastic volatility processes (Heston 1993), were intended from the outset to incorporate skewness and kurtosis. Hull (2003, p. 438) cited asymmetry in the log return distribution as an underlying cause of the well known implied volatility smirk observed in options markets. However, Savickas (2005) showed that models that allow skewness did not outperform the standard model in delta hedging tests on financial time series.

Given the ramifications, it is not surprising that authors have attempted to measure the asymmetry of returns distributions directly. The sophistication and methodology of these tests vary greatly, as do the results, with no consensus on whether asymmetry exists, or in which direction. Of course the results also depend on which kinds of data are analysed. The following are some attempts at a direct approach, with a bias towards recent results.

Kon (1984) used maximum likelihood to fit mixtures of normals and a Student's t to the 30 stocks in the Dow Jones Industrial Average and the S&P, EW, and VW indices using daily data from 2 July 1962 to 31 December 1980. He also calculated the classic skewness statistic, that is, the sample third central moment normalised by the square root of the cube of the sample variance for each time series. The finding was that skewness was significantly positive for all but four stocks, and significantly positive for two indices and significantly negative for one index, all at a one percent significance level. These significance levels were established under the assumption that returns were normally distributed, however, and so overestimate the significance of the skewness results, given that every time series used in the study exhibited significant leptokurtosis.

Peiró (1999) recognised that the significance of skewness results was drastically reduced under alternative null hypotheses that returns were Student's t distributed with various degrees of freedom. Peiró therefore used model free asymmetry statistics. These were computed by comparing the distributions of positive mean corrected returns against those of absolute negative mean corrected returns using Kolmogorov-

Smirnov, Wilcoxon rank-sum, and Siegel-Tukey tests. The test data incorporated daily data from 9 market indices and 3 exchange rates. The indices were the S&P 500 Composite, Dow Jones Industrial, Nikkei, Financial Times 100, Commerzbank, CAC General, Composite, Banca Commerciale Italiana, and General of the stock exchanges of New York (the first two), Tokyo, London, Frankfurt, Paris, Toronto, Milan and Madrid. The exchange rates were the Japanese Yen, British Pound, and the German Mark, all against the US Dollar. All of the data was drawn from the period 3 January 1980 to 27 September 1993, with the exception of the Financial Times 100, which was from 3 January 1984 to 27 September 1993. Peiró found that "in most markets daily financial returns are symmetric or, at least, do not present strong evidence of skewness". Some evidence of asymmetry was found in five of the time series, with three of them having a higher frequency of negative mean corrected returns.

Peiró went on to perform the same tests on daily, weekly, and monthly returns for 48 stocks from the New York Stock Exchange (Peiró 2002). The data was from the period 26 December 1995 to 25 May 2000 and included 24 Dow Jones stocks and 24 stocks with lower market capitalisation. The finding this time was that the tests on "daily returns detect some asymmetries, although asymmetry does not seem to be a stylized fact characteristic of daily returns", and that asymmetry disappeared as returns were aggregated into weekly and monthly data sets.

A different criticism of the classic skewness statistic was leveled in Kim and White (2004), where the point was made that it is not a robust statistic. Some robust statistics were therefore explored which compare quantiles and take values between -1 and 1. Monte Carlo simulations are used to show that the robust statistics were less sensitive to outliers, but no assertions were made about the distributions of the statistics. However, "all the robust skewness measures are pretty close to zero and hence indicate that there is little skewness" when tested on daily S&P500 index data from 1 Janurary 1982 to 29 June 2001.

Premaratne and Bera (2005) recognised both that the standard skewness statistic is inflated by leptokurtosis and that the third moment is not robust. A semiparametric test for asymmetry was attempted in this instance, with returns fitted to Pearson type IV distributions using maximum likelihood. As the authors commented, Pearson type IV distributions include the normal, Student's t, gamma, beta, and F distributions as special cases. The authors claimed that the test is semiparametric, because the test statistic was reduced to the expectation of the arctangent of the mean corrected returns. When tested on 29 stocks selected from the Center for Research in Security Price (CRSP) database, their statistic found that 24 stocks were significantly positively skewed. The standard statistic found the same number of significantly positively skewed stocks, although not the same 24, and some significantly negatively skewed stocks. The data used for the test was daily data from January 1990 to December 1996.

While all of the above direct tests have assumed explicitly or implicitly that returns are independently and identically distributed, Bai and Ng (2005) presented some tests for skewness and kurtosis assuming weakly dependent returns. This paper is interesting in that its skewness statistic was identical to the classic skewness statistic, but calculated the variance of this statistic under general assumptions, rather than assuming normality. These calculations assumed stationarity and take into account the higher moments and dependence structure of the time series, which can be measured empirically. Bai and Ng also calculated a skewness statistic that incorporated the third and fifth moments. For testing purposes they assumed the finiteness of the tenth moment. The empirical applications are mostly concerned with 21 American macroeconomic time series such as the quarterly GDP, monthly unemployment rate, and weekly 30-day interest rates over various time periods. However, three exchange rates were tested and two stock indices were tested. One exchange rate and both indices had significant negative skewness at a one percent significance level. The exchange rates were the Canadian Dollar, German Mark,

and Japanese Yen exchange rates against the US Dollar from January 1971 to December 1997 and the indices were the value and equally weighted CRSP daily stock returns from 2 January 1990 to 31 December 1996.

Although Ngatchou-Wandji (2006) did not present any empirical results, their paper is worth mentioning because it addressed similar skewness measures to those presented in Kim and White (2004), but provided appropriate normalisations in testing for significance. Power of the statistics was estimated using simulations and compared to that of the classic statistic. It assumed that returns are iid.

As stated in Section 6.1, the statistics proposed in Chapter 6 are valid in the face of the stylised features that characterise real daily log return samples, are nonparametric, and reduce to the standard skewness statistic if it is applicable. We recognise that, while the usual moment based estimators are not robust, in practice there are many instances where the third moment will be used in some way to measure skewness, so we maintain a link with the standard statistic to preserve the intuition associated with it. The next section gives our empirical contribution.

7.2 Empirical Results for the Generalised Skewness Statistic

The data used for the empirical test consisted of 614 time series taken from the Australian Stock Exchange (ASX). Each time series ended on 17 July 2006 and comprised at least 800 non-zero daily log returns of at most 15 years of data. The returns were adjusted for dividends and dilutions. The process by which data was acquired and cleaned is described in detail in Appendix 7.A. The analysis reported in this section uses the "RI" data described there.

As we are interested in whether the stock log returns display significant skewness, we study the $t_n(\beta)$ statistics for $1 < \beta < 3$. The $t_n(\beta)$ statistics, given in (6.2.8), are the studentised version of the generalised skewness statistics $g_n(\beta)$ defined in

(6.2.3), and should asymptotically follow a standard normal distribution for values of β sufficiently close to one by Theorem 6.2.3. The β range was chosen as it is the largest range that is interesting and may be permissible under the assumptions of Theorem 6.2.3. It also facilitates comparisons with the simulation results in Section 6.3.

Displaying the results of calculating $t_n(\beta)$ for 614 stocks and a continuous range of β represents a challenge not addressed in any of the studies mentioned in the last section, and a tabular display of results is not appropriate. Rather, Figure 7.1 shows a contour map of the empirical cumulative density function of $t_n(\beta)$ for varying β for the sample of 614 stocks. The contour lines show 5% quantiles from 0% to 100%. The dashed lines demarcate the rejection region for hypothesis H_0 in Section 6.2, that is that $g(\beta) = 0$, at a two-sided 5% significance level.

As expected from the simulation tests reported in Section 6.3, the quantiles of $t_n(\beta)$ shown in Figure 7.1 tend toward zero as β increases. It is interesting to compare the quantiles of $t_n(\beta)$ at $\beta = 3$, which correspond to the studentised version of the standard skewness statistic, with those of $t_n(\beta)$ at $\beta = 1.5$, which are more likely to be valid given the tail weight estimates of stock log return distributions described in Section 6.1. If we were to look only at $t_n(\beta)$ at $\beta = 3$, calculations show that we would underestimate the amount of skewness present and reject the null hypothesis of symmetry at a 5% significance level for less than 11% of the 614 stocks. This can be observed in Figure 7.1 on the right vertical axis, where the top three quantiles, corresponding to 100%, 95%, and 90% of the stocks, just clear the boundary of the upper rejection region, and the 0% quantile lies above the boundary of the lower rejection region. However, at $\beta = 1.5$, evidence for the asymmetry of the stock log returns is stronger, with calculations showing that the null hypothesis is rejected at a 5% significance level for more than 18% of the stocks. This can also be observed in Figure 7.1 as the contours arc upwards as one moves towards smaller values of β. Note that, in a two-sided test at a 5% significance level, the

only significantly asymmetric results in the test display positive values of $g_n(\beta)$.

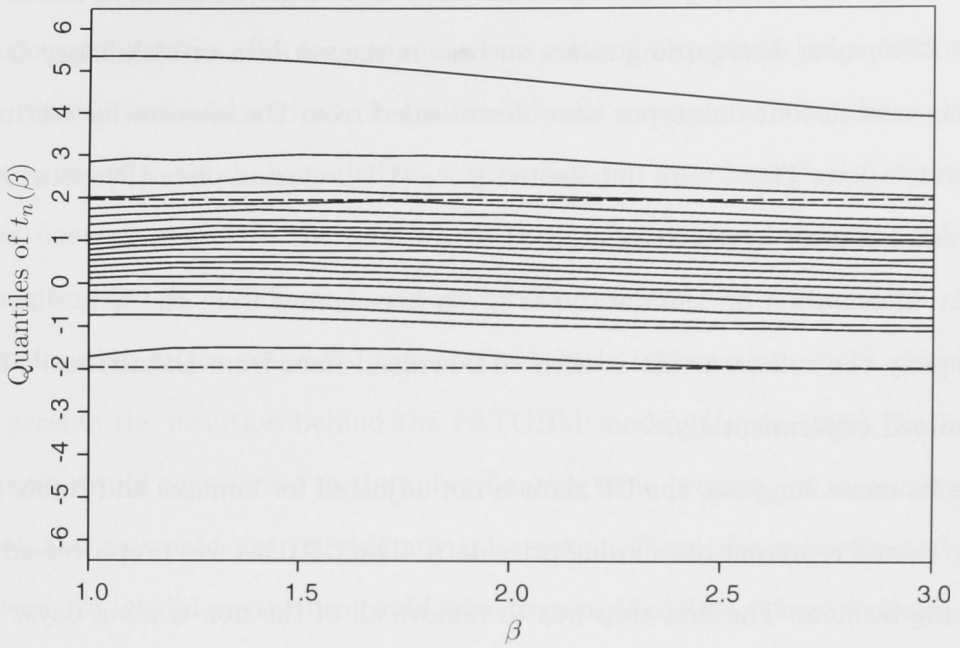

Figure 7.1: Contours show quantiles of $t_n(\beta)$ in 5% increments, from 0% to 100%, for varying β over 614 ASX stocks. Dashed lines show the two-sided rejection region at a 5% significance level.

Appendix 7.A Data Source and Data Cleaning

All of the data used in this study was downloaded from Thomson Financial's Datastream. This appendix describes the steps taken to preprocess the data. The market chosen for the study was the Australian Stock Exchange (ASX). In order to obtain enough data points in each time series to make the convergence results of Section 6.2 applicable, 15 years of daily data were requested, up to and including 17 July 2006, for each stock that was listed on Datastream as part of the ASX on that day. Before culling, this resulted in a dataset of 1762 series of 3798 return points, many of which were null for companies that had not traded over the entire interval. All data processing and testing was performed using S-PLUS.

Ideally, the data used for this study would consist of the logarithms of the ratios

of the accurate final closing prices of each stock on each day that stock was traded, adjusted for dividends and dilution effects. In reality it is difficult for any data vendor to track and apply all of the correction factors generated by a market such as the ASX, even disregarding issues such as miskeyed data or technology failures. For this reason, four datatypes were downloaded from Datastream for testing and cross-validation. These were unadjusted price (UP), closing price (P), return index (RI), and turnover by volume (VO).

The information on the datatypes given here comes from the help files which accompany the software that is used to download data from Datastream[1], or was determined experimentally.

As its name suggests, the UP data is not adjusted for bonuses and rights issues, and so should represent raw closing price data. The UP data was preprocessed in the following fashion. The first step was to remove all of the non-trading days. While the Datastream data did not contain weekend dates, it did contain approximately eight public holidays each year, with these days padded. Padding data is the practice of carrying price data forward where no data is available. The second step was to cull the data series with too few data points. So that the convergence results of Section 6.2 would be valid for each series, any stock that had less than 800 non-zero returns in the given time period was discarded, resulting in a set of 846 stocks. This raises two issues.

The first is survivorship bias. The original data set included 30 stocks that were classified by Datastream as dead or suspended at 17 July 2006. Of these 30, 12 remained in the set of 846 after culling. However, failed companies are still underrepresented in this set. For instance, Internet research revealed that more than 88 companies listed on the ASX were declared worthless between 1990 and March 2006. The criterion that at least 800 non-zero returns exist also excludes young stocks. Thus there is a possible survivorship bias that will be addressed

[1]Help file references are to the Definition help files under the Help submenu of the Datastream-AFO menu provided by the Datastream addin to Microsoft Excel.

again below.

The second issue was the overrepresentation of zero returns in many of the time series. It is for this reason that 800 non-zero returns were required rather than 800 non-null returns. To remove some zero returns requires a judgement of which zero returns are valid and which are invalid. An example of a valid zero return is where a stock is definitely traded on a given day, but has closing price exactly equal to its opening price. In the case where there is no trading of a particular stock on a given day, it is difficult to decide whether a zero return should be removed. This situation arises naturally, but provides little information about skewness. If one accepts the intuition behind the FATGBM model (Heyde 1999), for instance, this case would correspond to the subordinating time process not increasing during a day, and so would not provide any information about the subordinated Wiener process. An invalid zero return could arise from padded data, where a closing price is carried forward in the absence of new data, possibly as a result of a genuine error, for example from a technology failure, or where trading is suspended or a stock delisted. Invalid zeros can also be caused by other genuine errors, for instance where a series of zeros is erroneously entered in place of a series of nulls in the original price series.

For our purposes, the only zero returns that were judged to be valid were those where trading definitely occurred but closing price equalled opening price. The turnover by volume (VO) data was downloaded to identify such cases. The VO data gives the number of shares traded on a given day in units of thousands of shares. One implication of this is that if less than a thousand but more than zero shares were traded for a stock on a given day, no volume would have been registered. Another peculiarity of this data set is that zero volume is usually recorded as null volume.

The UP data was filtered using the VO data. Any zero return observed in the UP data that coincided with a zero or null data point in the VO data was set to null. On days when less than a thousand but more than zero shares were traded and closing price equalled opening price, this procedure resulted in zero returns being

erroneously removed, but there were few such cases and this resulted in less errors than leaving all of the zero returns in the UP data. In all, approximately 49% of the zero returns present in the UP data were removed in this way.

The Datastream datatype definition files describe P data "prices taken at the close of market" that are "adjusted for subsequent capital actions". The P data was filtered in exactly the same fashion as the UP data. Public holidays were removed, the same series were selected, and VO filtering was applied. One possible problem with the P data that did not exist for the UP data is that more errors may have been introduced by Datastream when the prices were adjusted for capital actions. Consider the hypothetical case where a dilution caused the UP price to fall by half on some day. If the adjustment was applied on the wrong day, then two return data points will be incorrect: the day of the dilution, and the day on which the adjustment was erroneously applied. There may also be series where Datastream has missed adjustments altogether.

The RI datatype is described in the help files as closing prices adjusted for capital actions and dividends. Initial filtering for the RI data followed the same procedure as for the UP and P data with regard to public holidays and time series selection. However, more filtering was then required. The method that is used to calculate RI data is given in some detail in its datatype definition. Again according to the help files, the method by which the returns are adjusted for dividends in the Australian market changed in 1988. Before this time, the following formula was used:

$$RI_t = RI_{t-1} \times \frac{PI_t}{PI_{t-1}} \times \left(1 + \frac{DY_t}{100} \times \frac{1}{N}\right),$$

where t and $t-1$ refer to day t and the previous day, RI is the data series, PI is the "price index" which is presumably the P data, DY is the dividend yield as a percentage, and N is the number of working days per year, taken to be 260. After

1988, the formula for incorporating dividend returns was changed to

$$RI_t = \begin{cases} RI_{t-1} \times \frac{P_t}{P_{t-1}}, & t \text{ is not a dividend ex-date,} \\ RI_{t-1} \times \frac{P_t + D_t}{P_{t-1}}, & t \text{ is a divdend ex-date,} \end{cases}$$

where D is the dividend amount.

This level of explanation was useful in understanding some undisclosed features of the RI data. For some series for which dividends were known to have occurred from other sources, such as company websites, sharp differences in returns between the RI and P data series were not observed on dividend ex-dates. In such cases, when the time series were viewed over the time frame of approximately a year, RI returns were found to be systematically higher than P returns. The existence of rounding meant that this difference was variable, but on average positive. Given this, one can speculate that the pre-1988 formula is used as a default where Datastream has, for whatever reason, missed dividends for a particular company. This means that an additional artificial drift term is added to some RI data series in some years, destroying any stationarity assumption. Therefore, a filtering rule was used to sort the "good" RI series, where dividends were incorporated correctly, from "bad" RI series, where they were not.

The rule itself is heuristic. For good RI series, the difference between the RI and P series is small and symmetric around zero, except in a handful of cases where dividends occured. For bad RI series, the difference between the RI and P series was still small, but usually positive. Based on this observation, the following quantity was calculated for each series:

$$R = \max_{i \in \{1,\ldots,n-125\}} \sum_{j=0}^{125} \operatorname{sgn}\left(X^{\mathrm{RI}}_{i+j} - X^{\mathrm{P}}_{i+j}\right).$$

A period of 126 returns (six months) was chosen as it is unlikely that Datastream review their dividends estimates more frequently than every six months. The rejec-

tion rule is then: $R < 30$ implies a good series; $R \geq 30$ implies a bad series. This seemingly arbitrary rule is justified in Figure 7.2, which shows a histogram of the results over the set of 846 series. Note that the P data that was used for comparison with the RI data was not filtered using the VO data. The set of good RI returns series was reduced in this way to 614 time series.

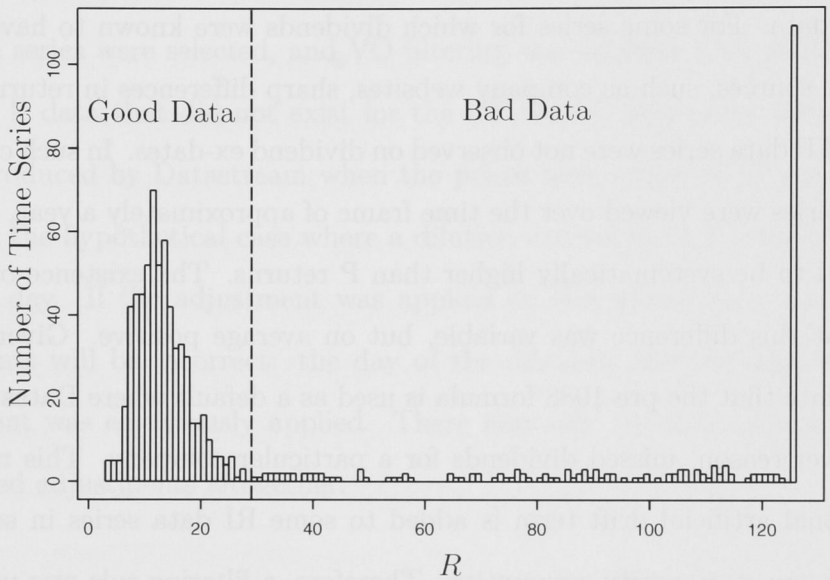

Figure 7.2: Histogram of R Values for 846 RI Time Series

The reduced RI data set was then filtered to reduce the number of zero returns using the VO data in the same manner that the UP and P data were filtered. The filtered RI data set was used to generate the results Section 7.2.

Cleaning the data did not alter or remove any non-zero data points, and was completely independent of the subsequent statistical tests. Returning to the question of survivorship bias: as a rough test, the simple statistic of log market cap was compared between included stocks and all available stocks using a QQ-plot. No bias was evident in this comparison, so we assume it not present in the RI data either.

Chapter 8

Conclusion

Our results allow the significance of the asymmetry of a time series of leptokurtic stationary ergodic martingale differences to be measured. The proposed statistic is a generalisation of the usual skewness statistic that is based on the sample third moment, and reduces to that statistic if it is applicable.

The new statistic was applied to a large sample of stocks from the Australian Stock Exchange. Whereas the usual skewness statistic was found to comparatively underestimate the presence of skewness in the market, the new statistic showed more rigorously that greater than 18% of the stocks surveyed displayed significant positive skewness, and that negative skewness was not significant for any of the sample stocks. All tests were two-sided at the 5% significance level.

Bibliography

Abramowitz, M. and I. A. Stegun (1964). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables (tenth GPO printing ed.). New York: Dover.

Almendral, A. (2005). Numerical valuation of American options under the CGMY process. In A. E. Kyprianou, W. Schoutens, and P. Wilmott (Eds.), Option Pricing and Advanced Lévy Models, pp. 259–276. Chichester: Wiley.

Almendral, A. and C. W. Oosterlee (2007). On American options under the variance gamma process. Appl. Math. Finance 14(2), 131–152.

Amin, K. I. (1993). Jump diffusion option valuation in discrete time. J. Finance 48(5), 1833–63.

Andricopoulos, A. D., M. Widdicks, P. W. Duck, and D. P. Newton (2003). Universal option valuation using quadrature methods. Journal of Financial Economics 67(3), 447–471.

Bai, J. and S. Ng (2005). Tests for skewness, kurtosis, and normality for time series data. J. Bus. Econom. Statist. 23(1), 49–60.

Barndoff-Nielsen, O. E., J. Pedersen, and K.-I. Sato (2001). Multivariate subordination, self-decomposability and stability. Adv. in Appl. Probab. 33, 160–187.

Barndorff-Nielsen, O. E. (1997). Normal inverse gaussian distributions and stochastic volatility modelling. Scand. J. Statist. 24(1), 1–13.

Barron, A. R. (1993, May). Universal approximation bounds for superpositions of a sigmoidal function. IEEE Trans. Information Theory 39(3), 930–945.

Billingsley, P. (1995). Probability and Measure (3rd ed.). New York: John Wiley.

Billingsley, P. (1999). Convergence of Probability Measures (2nd ed.). New York: John Wiley.

Bingham, N. H., R. Kiesel, and R. Schmidt (2003). A semi-parametric approach to risk management. Quant. Finance 3(6), 426–441.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. J. Pol. Econ. 81(3), 637–654.

Bouchard, B., N. E. Karoui, and N. Touzi (2005). Maturity randomization for stochastic control problems. Ann. Appl. Probab. 15(4), 2575–2605.

Boyarchenko, M. (2008, September). Carr's randomization for finite-lived barrier options: Proof of convergence. http://ssrn.com/abstract=1275666.

Boyarchenko, S. I. and S. Z. Levendorskiĭ (2002). Perpetual american options under lévy processes. SIAM J. Control Optim. 40(6), 1663–1696.

Breiman, L. (1993, May). Hinging hyperplanes for regression, classification, and function approximation. IEEE Trans. Information Theory 39(3), 999–1013.

Briys, E., M. Bellalah, H. M. Mai, and F. de Varenne (1998). Options, Futures, and Exotic Derivatives: Theory, Application and Practice. Chichester: John Wiley & Sons.

Broadie, M. and J. B. Detemple (2004). Option pricing: Valuation models and applications. Management Sci. 50(9), 1145–1177.

Caballero, M., J. Pardo, and J. Pérez (2008). On the lamperti stable processes. arXiv:0802.0851v3.

Carr, P. (1998). Randomization and the american put. Rev. Finan. Stud. 11(3), 597–626.

Carr, P., H. Geman, D. B. Madan, and M. Yor (2002). The fine structure of asset returns: An empirical inestigation. J. Bus. 75(2), 305–332.

Carr, P. and A. Hirsa (2003). Why be backward? forward equations for American options. Risk 16(1), 103–107.

Chang, C.-C., S.-L. Chung, and R. C. Stapleton (2007). Richardson extrapolation techniques for the pricing of american-style options. J. Futures Markets 27(8), 791–817.

Clift, S. S. and P. A. Forsyth (2008). Numerical solution of two asset jump diffusion models for option valuation. Appl. Numer. Math. 58(6), 743–782.

Cont, R. and P. Tankov (2004). Financial Modelling with Jump Processes. Boca Raton, Florida: Chapman and Hall/CRC.

Coquet, F., J. Mémin, and L. Słominski (2001). On weak convergence of filtrations. In M. Émery and P.-A. Meyer (Eds.), Séminaire de Probabilités de Strasbourg XXXV, Volume 1755 of Lecture Notes in Mathematics, pp. 306–328. Berlin: Springer.

Coquet, F. and S. Toldo (2007). Convergence of values in optimal stopping and convergence of optimal stopping times. Electron. J. Probab. 12, 207–228.

Cox, J. C., S. A. Ross, and M. Rubinstein (1979, September). Option pricing: A simplified approach. J. Finan. Econ. 7(3), 229–263.

Daníelsson, J. and C. G. de Vries (1997). Tail index and quantile estimation with very high frequency data. J. Empirical Finance 4, 241–257.

Daníelsson, J. and C. G. de Vries (1998, February). Value-at-risk and extreme returns. Tinbergen Institute Discussion Papers 98-017/2, Tinbergen Institute. available at http://ideas.repec.org/p/dgr/uvatin/19980017.html.

Derman, E. (1996). Outperformance options. In I. Nelkin (Ed.), The Handbook of Exotic Options: Instruments, Analysis and Applications, Chapter 9, pp. 244–250. McGraw-Hill.

DeVore, R. (1998). Nonlinear approximation. Acta Numer. 7, 51–150.

d'Halluin, Y., P. Forsyth, and G. Labahn (2004). A penalty method for American options with jump diffusion processes. Numer. Math. 97(2), 321–352.

Eberlein, E., K. Glau, and A. Papapantoleon (2008, September). Analysis of valuation formulae and applications to exotic options in Lévy models. arXiv:0809.3405v1.

Eberlein, E. and A. Papapantoleon (2005). Symmetries and pricing of exotic options in Lévy models. In A. Kyprianou, W. Schoutens, and P. Wilmott (Eds.), Exotic Option Pricing and Advanced Lévy Models, pp. 99–128. Wiley.

Friedman, J. H. (1991a). Multivariate adaptive regression splines. Ann. Statist. 19(1), 1–67.

Friedman, J. H. (1991b). Rejoinder: Multivariate adaptive regression splines. Ann. Statist. 19(1), 123–141.

Galassi, M., J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth, and F. Rossi (2008). Gnu Scientific Library: Reference Manual (1.12 ed.). Network Theory Ltd.

Gamma, E., R. Helm, R. Johnson, and J. Vlissides (1995). Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley Professional.

Geske, R. and H. E. Johnson (1984). The american put valued analytically. J. Finance 39(5), 1511–1542.

Harvey, C. R., J. Liechty, M. W. Liechty, and P. Mueller (2004). Portfolio selection with higher moments. available at http://citeseer.ist.psu.edu/harvey04portfolio.html.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Finan. Stud. 6, 327–343.

Heyde, C. C. (1999). A risky asset model with strong dependence through fractal activity time. J. Appl. Probab. 36, 1234–1239.

Heyde, C. C. and P. Hall (1980). Martingale Limit Theory and its Application. New York: Academic Press.

Heyde, C. C. and N. N. Leonenko (2005, February). Student processes. Adv. in Appl. Probab. 37(2), 342–365.

Heyde, C. C. and S. Liu (2001). Empirical realities for a minimal description risky asset model. the need for fractal features. J. Korean Math. Soc. 38(5), 1047–1059.

Howison, S. (2007). A matched asymptotic expansions approach to continuity corrections for discretely sampled options. part 2: Bermudan options. Appl. Math. Finance 14(1), 91–104.

Hull, J. C. (2003). Options, Futures and Other Derivatives (4th ed.). Prentice Hall.

Hunziker, J. P. and P. Koch-Medina (1996). Two-color rainbow options. In I. Nelkin (Ed.), The Handbook of Exotic Options: Instruments, Analysis and Applications, Chapter 5, pp. 143–174. McGraw-Hill.

Hurst, S. R. and E. Platen (1997). The marginal distribution of returns and volatility. In Y. Dodge (Ed.), L_1-Statistical Procedures and Related Topics, Volume 31 of IMS Lecture Notes-Monograph Series, pp. 301–314. Hayward, CA: IMS.

Hurst, S. R., E. Platen, and S. T. Rachev (1997). Subordinated market index models: A comparison. Finan. Eng. Japanese Markets 4, 97–124.

Jacod, J. and A. N. Shiryaev (2002). Limit Theorems for Stochastic Processes (2nd ed.). Verlag Berlin Heidelberg: Springer.

Jansen, D. W. and C. G. de Vries (1991). On the frequency of large stock returns: Putting booms and busts into perspective. Rev. Econom. Statist. 73(1), 18–24.

Jones, L. K. (1992). A simple lemma on greedy approximation in hilbert space and convergence rates for projection pursuit regression and neural network training. Ann. Statist. 20(1), 608–613.

Karatzas, I. and S. E. Shreve (1991). Brownian Motion and Stochastic Calculus (2nd ed.). Verlag New York Berlin Heidelberg: Springer.

Kargin, V. (2005). Lattice option pricing by multidimensional interpolation. Math. Finance 15(4), 635–647.

Këllezi, E. and N. Webber (2004). Valuing bermudan options when asset returns are Lévy processes. Quant. Finance 4(1), 87–100.

Kendall, M. G., A. Stuart, and J. K. Ord (1994). Kendall's Advanced Theory of Statistics (6th ed.), Volume 1: Distribution Theory. London, United Kingdom: Arnold.

Kim, T.-H. and H. White (2004, September). On more robust estimation of skewness and kurtosis: Simulation and application to the S&P500 index. University of California at San Diego, Economics Working Paper Series 2003-12, Department of Economics, UC San Diego. available at http://ideas.repec.org/p/cdl/ucsdec/2003-12.html.

Kon, S. J. (1984). Models of stock returns — a comparison. J. Finance 39(1), 147–165.

Konno, H., H. Shirakawa, and H. Yamazaki (1993). A mean-absolute deviation-skewness portfolio optimization model. Ann. Oper. Res. 45(1), 205–220.

Korkie, B., R. Sivakumar, and H. J. Turtle (2006, February). Variance spillover and skewness in financial asset returns. The Financial Review 41(1), 139–156. available at http://ideas.repec.org/a/bla/finrev/v41y2006i1p139-156.html.

Levendorskiĭ, S. Z. (2004). Pricing of the American put under Lévy processes. Int. J. Theor. Appl. Finance 7(3), 303–336.

Longstaff, F. A. and E. S. Schwartz (2001). Valuing American options by simulation: A simple least-squares approach. Rev. Finan. Stud. 14(1), 113–147.

Lord, R., F. Fang, F. Bervoets, and C. W. Oosterlee (2007). A fast method for pricing early-exercise options with the fft. In Computational Science — ICCS 2007, Volume 4488 of Lecture Notes in Computer Science, pp. 415–422. Berlin / Heidelberg: Springer.

Madan, D. B., P. P. Carr, and E. C. Chang (1998). The variance gamma process and option pricing. Europ. Finance Rev. 2, 79–105.

Madan, D. B. and E. Seneta (1990). The variance gamma (V.G.) model for share market returns. J. Bus. 63(4), 511–524.

Maller, R. A., D. H. Solomon, and A. Szimayer (2006). A multinomial approximation for American option prices in Lévy process models. Math. Finance 16(4), 613–633.

Martzoukos, S. H. (2003). Contingent claims on foreign assets following jump-diffusion processes. Rev. Derivatives Res. 6(1), 27–45.

Matache, A.-M., P.-A. Nitsche, and C. Schwab (2005). Wavelet galerkin pricing of American options on Lévy driven assets. Quant. Finance 5(4), 403–424.

Matache, A.-M., C. Schwab, and T. P. Wihler (2005). Fast numerical solution of parabolic integrodifferential equations with applications in finance. SIAM J. Sci. Comput. 27(2), 369–393.

Meyer, P. A. and W. A. Zheng (1984). Tightness criteria for laws of semimartingales. Ann. Inst. H. Poincaré Probab. Statist. 20, 353–372.

Mordecki, E. (2002). Optimal stopping and perpetual options for lévy processes. Finance Stoch. 6(4), 473–493.

Mulinacci, S. and M. Pratelli (1998). Functional convergence of snell envelopes: Applications to American options approximations. Finance Stoch. 2(3), 311–327.

Ngatchou-Wandji, J. (2006). On testing for the nullity of some skewness coefficients. Internat. Statist. Rev. 74(1), 47–65.

Ong, M. (1996). Exotic options: The market and their taxonomy. In I. Nelkin (Ed.), The Handbook of Exotic Options: Instruments, Analysis and Applications, Chapter 1, pp. 3–44. McGraw-Hill.

Patton, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. J. Finan. Econ. 2(1), 130–168.

Peiró, A. (1999). Skewness in financial returns. J. Banking Finance 23, 847–862.

Peiró, A. (2002). Skewness in individual stocks at different investment horizons. Quant. Finance 2(2), 139–146.

Petrov, V. V. (1995). Limit Theorems of Probability Theory, Volume 4 of Oxford Studies in Probability. Oxford: Oxford Science Publications.

Pham, H. (1997). Optimal stopping, free boundary, and American option in a jump-diffusion model. Appl. Math. Optim. 35, 145–164.

Platen, E. and R. Sidorowicz (2008). Empirical evidence on student-t log-returns of diversified world indices. J. Statist. Theory and Practice 2(2), 233–251.

Premaratne, G. and A. Bera (2005). A test for symmetry with leptokurtic financial data. J. Finan. Economet. 3(2), 169–187.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992). Numerical Recipes in C (2nd ed.). Cambridge University Press.

Prigent, J.-L. (2003). Weak Convergence of Financial Markets. Berlin Heidelberg: Springer-Verlag.

Rydberg, T. H. (2000). Realistic statistical modelling of financial data. Internat. Statist. Rev. 68, 233–258.

Sato, K. (1999). Lévy Processes and Infinitely Divisible Distributions. Cambridge University Press.

Savickas, R. (2005, June). Evidence on delta hedging and implied volatilities for the black-scholes, gamma, and wiebull option pricing models. J. Finan. Res. 28(2), 299–317.

Schoutens, W. (2006). Exotic options under lévy models: an overview. J. Comput. Appl. Math. 189(1), 526–538.

Semeraro, P. (2006). A multivariate time-changed Lévy model for financial applications. Working Paper 10/2006, International Center for Economic Research.

Semeraro, P. (2008). A multivariate variance gamma model for financial applications. Int. J. Theor. Appl. Finance 11(1), 1–18.

Sengul, B. (2008). Lamperti stable processes for modelling of financial data. Technical report, University of Bath.

Szimayer, A. and R. A. Maller (2007). Finite approximation schemes for Lévy processes, and their application to optimal stopping problems. Stochastic Process. Appl. 117(10), 1422–1447.

Temlyakov, V. (2003). Nonlinear methods of approximation. Found. Comput. Math. 3, 33–107.

Villeneuve, S. (1999). Exercise regions of american options on several assets. Finance Stoch. 3(3), 295–322.

Wong, D. (1996). Generalised Optimal Stopping Problems and Financial Markets. Essex: Addison Wesley Longman.